DOCUMENT RESUME

ED 158 231 CS 004 273

AUTHOR	Samuels, S. Jay, Ed.
TITLE	What Research Has to Say about Reading Instruction.
INSTITUTION	International Reading Association, Newark, Del.
PUB DATE	78
NOTE	186p.; A number of figures were removed due to copyright restrictions
AVAILABLE FROM	International Reading Association, 800 Barksdale Road., Newark, Delaware 19711 (Order No. 488, $4.25 member, $6.25 non-member)
EDRS PRICE	MF-$0.83 HC-$10.03 Plus Postage.
DESCRIPTORS	Criterion Referenced Tests; Educational Assessment; Learning Disabilities; Linguistics; Psycholinguistics; Reading Ability; Reading Comprehension; *Reading Difficulty; Reading Improvement; *Reading Instruction; Reading Programs; Reading Rate; *Reading Research; *Research Utilization; *Theories
IDENTIFIERS	Piaget (Jean)

ABSTRACT

The papers in this volume represent both theory and application in reading research. The topics are as follows: research on programs for poor readers; a method for teaching children who have reading difficulties even though they have learned decoding skills; a decade of research on reading and learning disabilities; children's understanding of linguistic units; reading research that assists classroom instruction; flexibility of reading rate and efficient reading behavior; applications of a psycholinguistic model of reading; implications of Piaget's theory for reading instruction; comprehension abilities; the role of content in teaching reading; assessment; criterion referenced measurement in a norm referenced context; and a pragmatic-empirical approach to research in reading. Each entry is preceded by a brief introduction, and tables and bibliographies accompany many of the papers. (MAI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED158231

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

What Research Has to Say about Reading Instruction

S. Jay Samuels, Editor
University of Minnesota

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

International Reading Association

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

CS504223

INTERNATIONAL READING ASSOCIATION
800 Barksdale Road
Newark, Delaware 19711

INTERNATIONAL READING ASSOCIATION

OFFICERS
1977-1978

President William Eller, State University of New York at Buffalo, Amherst, New York

Vice President Dorothy S. Strickland, Kean College of New Jersey, Union, New Jersey

Vice President-Elect Roger Farr, Indiana University, Bloomington, Indiana

Executive Director Ralph C. Staiger, International Reading Association, Newark, Delaware

DIRECTORS

Term expiring Spring 1978

Roselmina Indrisano, Boston University, Boston, Massachusetts
Ethna R. Reid, Exemplary Center for Reading Instruction, Salt Lake City, Utah
Robert B. Ruddell, University of California, Berkeley, California

Term expiring Spring 1979

Lou E. Burmeister, University of Texas, El Paso, Texas
Jack Cassidy, Newark School District, Newark, Delaware
Kenneth S. Goodman, University of Arizona, Tucson, Arizona

Term expiring Spring 1980

Mary Ann Baird, State Department of Education, Jackson, Mississippi
Addie Stabler Mitchell, Morehouse College, Atlanta, Georgia
Jean E. Robertson, University of Alberta, Edmonton, Alberta

Copyright 1978 by the
International Reading Association, Inc.

Library of Congress Cataloging in Publication Data
Main entry under title:
What research has to say about reading instruction.
Bibliography: p.
1. Reading—Addresses, essays, lectures. 2. Reading research—Addresses, essays, lectures. I. Samuels, S. Jay.
LB1050.W435 428'.4 78-4573
ISBN 0-87207-488-0

Contents

Foreword *v*

Preface *vi*

1 *Clues from Research on Programs for Poor Readers* John T. Guthrie, Mary Seifert, and Lloyd W. Kline

13 *When You Still Can't Read in Third Grade: After Decoding, What?* Carol Chomsky

31 *A Decade of Research on Reading and Learning Disabilities* Jeanne Chall

43 *Children's Understanding of Linguistic Units* Walter H. MacGintie

57 *Research in Reading that Should Make a Difference in Classroom Instruction* Harry Singer

72 *Considering Flexibility of Reading Rate for Assessment and Development of Efficient Reading Behavior* Phyllis A. Miller

84 *Some Practical Applications of a Psycholinguistic Model of Reading* P. David Pearson

98 *Implications of Piaget's Theory for Reading Instruction* Frank B. Murray

109 *Developing Comprehension Abilities: Implications from Research for an Instructional Framework* Robert B. Ruddell

121 *The Role of Content in the Teaching of Reading* Gaston E. Blom

135 *What Research Can Tell the Reading Teacher about Assessment* Robert C. Calfee, Priscilla A. Drum, and Richard D. Arnold

165 *Criterion Referenced Measurement in a Norm Referenced Context* J. Jaap Tuinman

174 *A Pragmatic-Empirical Approach to Research in Reading* Wayne Otto

IRA PUBLICATIONS COMMITTEE 1977-1978 *Chairing*, Harold L. Herber, Syracuse University • Janet R. Binkley, IRA • Faye R. Branca, IRA • Roger Farr, Indiana University • Mary S. Feely, IRA • Lynette Saine Gaines, University of South Alabama • Margaret Keyser Hill, Southern Illinois University • Laura S. Johnson, Morton Grove, Illinois • Lloyd W. Kline, IRA • Connie Mackey, Montreal, Quebec • John McInness, Ontario Institute for Studies in Education, Toronto • John E. Merritt, Open University • Theodore A. Mork, Western Washington State College • Clifford D. Pennock, University of British Columbia • Emma W. Rembert, Florida International University • Robert B. Ruddell, University of California at Berkeley • Cyrus F. Smith, Jr., University of Wisconsin at Milwaukee • Zelda Smith, Gables Academies, Miami • Ralph C. Staiger, IRA • Carl Wallen, Arizona State University • Sam Weir.traub, State University of New York at Buffalo.

The International Reading Association attempts, through its publications, to provide a forum for a wide spectrum of opinion on reading. This policy permits divergent viewpoints without assuming the endorsement of the Association.

Foreword

Primitive Man was driven by deep desires to extend power beyond his physical and mental limits, to be in several places at once, to know the reality beyond his vision, to anticipate the future and prepare for it. Life for him was not the same every day. Its hazards were always around him and ahead of him, and he knew it.

So it is in the field of reading. It is a changing and moving field. As new evidence and new theories crowd into it, our preparation for professional roles, our production and use of materials and techniques, must reflect them. New developments demand that we modify our attitudes, programs, and convictions even after many years in the field.

During the 1975 Convention of IRA we could not possibly have attended all the meetings in which the papers in this volume were presented. Nor could we have recognized their interrelationships so readily as we can now. They answer in many ways the question: What should we be doing now, as professional people, to incorporate these ideas into our operations?

It is easy to be impulsive about change: either to refuse to consider it, or to discard something of value for something of less or different value. Neither reaction is the way to progress.

Progress will come through a change in ourselves as we read this volume for a better sense of values, a broader understanding of what we thought we knew. And this, of course, can only take place if we are willing to accommodate new ideas, as when climbing to a higher elevation we see in the landscape the relationships we had not sensed before.

S. Jay Samuels of the University of Minnesota has devoted himself most generously to the purposes and activities of the International Reading Association for a number of years. This is one of the latest of his many contributions. He not only planned the research sequence from which these papers are drawn, but now has compiled them in this convenient form.

The next step rests with us, the readers who, in our individual roles in whatever part of the world, will assess these ideas and put them to work. Then we may be worthy of the man who was driven by deep desires long ago, and may have a future, too.

CONSTANCE M. MCCULLOUGH

Preface

This volume is an outgrowth of an idea which Constance McCullough had when she was President-Elect of the International Reading Association. Having observed the annual convention for many years, she was of the opinion that too many papers given at the convention were either overly prescriptive—teachers were told what to do without explaining the rationale—or they were overly theoretical—a rationale was given without explaining in sufficient detail what were the practical implications. As organizer of IRA's 1975 annual convention, it was Dr. McCullough's idea to devote prime program time to a series of presentations which would balance theory and application. As Chairperson of the IRA Studies and Research Committee, it was my responsibility to invite the participants. Judging from the heavy attendance at the presentations and the positive audience response, Dr. McCullough's idea was highly successful.

This volume was compiled in order to preserve the many excellent papers given at the convention and to share them with a larger audience. In addition to the original research application papers given at the 1975 convention, the volume includes invited papers prepared by a group of distinguished researchers widely known for their contributions to reading. How well this volume meets the goals of combining theory and application only the reader can determine.

A note of appreciation should be extended to the contributors to this volume; to Harry Singer, for suggesting that we publish the papers; to Patricia Dahl, who helped in the organization and planning of the meeting; to Mary Ann Pittman and Ellen Sylvester, for preparing a number of the introductions to the papers; and to Rosalind Horowitz, for the editing done on this volume.

S. JAY SAMUELS

Reading instruction may be characterized as a three part process, with goal setting, instruction, and evaluation as its components. As the teacher evaluates the products of instruction, it may be discovered that some children require remedial instruction. Does remedial instruction help the child? The authors of this article have searched the research literature in order to answer this question as well as others: Do remedial programs carry over? What are the characteristics of successful programs? Is remedial instruction most effective with younger or older children? What is the most desirable pupil/teacher ratio? One additional question the reader of this article may wish to consider is "How is remedial reading instruction different from regular instruction?"

Clues from research on programs for poor readers

John T. Guthrie
Mary Seifert
Lloyd W. Kline
International Reading Association

The rationale by which remedial reading programs have been justified seems logical enough. Most youngsters learn to read to a certain level of ability within a certain period of time—to first grade level by the end of one year of schooling, to second grade level by the end of two years of schooling, and so forth. Some youngsters, however, fall behind the norm in their achievement. Something fails to work according to schedule for them. This one reads only on a first grade level by the end of third grade; that one is at preprimer after a full year of instruction; another is unable to read even one word after two years of schooling. What must be done in our schools for those labeled "poor readers," those who fall measurably behind the norm? The obvious logical answer: remediate.

Does that logical rationale describe a valid approach to instruction in reading? Does logic lead to truth? In blunter terms, does remedial reading work? If so, which methods work best for which poor readers? Which materials, which instructional techniques, which patterns of pupil assignment, which settings, what kinds of teachers? How much remediation is needed for which learners? How often should their classes meet? For how long at a time? How should we define the "norm" in pupil achievement? How can we identify a "poor reader"? By what criteria?

Then, the biggest question of all for those of us who insist that there should be some mutual benefits in the relationships between

This paper was partially supported by The Spencer Foundation.

practitioners and researchers: Can research answer any of these obvious but critical questions?

In another paper, "Reading Comprehension Processes and Instruction," Guthrie (1977) lists three potential uses that implications of research in reading might offer to practitioners. Implications of research would qualify as directly significant in the teaching of reading, if they "1) suggested bold new practices, 2) enabled us to select the most promising among an array of prevailing practices, or 3) served as an explanation of why certain educational activities are effective" (p. 293). Innovation, selection, explanation—at least these three are quite likely among the broad functions by which research can serve the practitioner.

Which of these three uses of research is most pertinent to a particular teacher of reading in a particular school situation? *That* question must be answered not by the researcher so much as by *that* teacher. Most often, research says simply, "This is or this happened under these specified conditions." Like the star in one of Frost's poems, research results say only "I burn." They do not necessarily tell us that Teacher X, given similar conditions, will achieve the same results in Classroom Y with student Z during the next school term. Neither can research decide for Teacher X whether the implications of the research being reported would represent an innovation, a selection process, or an explanation for that teacher in Classroom Y with Student Z. At best, we can say that good research will offer the practitioner a report of some certainty. We can hope that the interpretation of that report will provide an improvement over trial and error as a basis for planning programs.

Searching what has been reported

What, then, does research say to the question of special instruction for poor readers?

First, it is obvious to anyone at all familiar with special reading instruction that programs for poor readers come in a wide variety of sizes and shapes. The "teachers" in those programs range from fellow classmates to intensively trained reading specialists. The youngsters identified as poor readers in our schools vary widely in abilities and age, and they are "taught" in a range of teacher-pupil ratios. Instructional materials and techniques vary almost from school to school, teacher to teacher, and in many situations from learner to learner. Second, of the many reports of efforts to improve the reading levels of poor readers, few provide sufficient information by which to judge the impact of whichever program on those children.

In an effort to focus as specifically and meaningfully as possible on our basic question, what does research say about special instructional programs for poor readers, our review of the research

literature was carried out within certain limits. It required that the reports concern learners between the ages of seven and fifteen years, that reading tests had been given to those students both at the beginnings and ends of their instructional programs, and that the total durations of programs for those students had been reported in hours or weeks. Additional information was available, obviously, from many of the reports selected within those limitations. Of the research reports cited in *Education Index* and *Psychological Abstracts* from 1969 to 1975, fifteen studies met the criteria named here.

For each of those fifteen studies, learning rates were computed for all students. A learning rate in reading is a ratio that describes how rapidly the child has acquired reading ability in relation to the months of instructional time he has been given in reading. To calculate the ratio, the child's gain in grade equivalent (in months) on a reading test is divided by the length of time (in months) required to produce that gain. For example, suppose a normal child were beginning fourth grade and his grade equivalent in reading were 2.5. For purposes of calculation, he had been in school for thirty months but had acquired only fifteen months of reading ability. (Note that a 2.5 grade equivalent represents a gain of only 1.5 years, since a child beginning first grade is given a score of 1.0 grade level. Note also that most standardized tests are normed on a ten month basis, assuming no gain in reading ability during July and August.) In our example, the child's ratio would be fifteen months' reading gain divided by thirty months in school, or a learning rate of .50 in reading. He would have acquired reading ability about half as quickly as an average reader of the same age. A negative learning rate can result if a child's score on a reading test decreases with the passage of time.

We compared learning rates of the pupils in these fifteen studies before, during, and after their participation in remedial programs. It is around these learning rates that other elements of the fifteen studies are here compared and contrasted.

Do remedial programs carry over?

Before trying to identify those characteristics of remedial programs that seem to make a difference in learning rates, it is prudent to ask if remedial programs in fact substantially improve learning to read at all.

At least two of the fifteen studies have rigorously compared the reading achievement of groups of children who received remedial instruction with control groups of children who received normal classroom instruction. (An important feature of rigorous comparison is random assignment of students to treatment and control groups.) In two studies (Shaver and Nuhn, 1971; Strang, 1974), middle-class children who were reading one or more years

behind their expected levels, determined from their mental ages, were randomly divided into treatment and control groups. Their grade placements varied from fourth to tenth. Instruction consisted of tutoring with teacher-pupil ratios of 1:1 to 1:3. In the Strang study, audiovisual equipment was used extensively, and the program lasted for fifty-four hours at four hours of instruction per week. In the Shaver and Nuhn study, trained adult tutors worked with children for about one hundred seventy-five hours at five hours per week, using basal readers and phonic workbooks.

In both studies, the tutored groups scored significantly higher on standardized reading tests than the nontutored control groups at the end of the program, although the groups had been nearly identical (not significantly different) at the outset of remediation. Children from all grades—fourth, sixth, seventh, and tenth—showed the effect. The measured increase in reading level for the tutored groups over that for nontutored groups cannot be attributed to statistical regression because both control and treatment groups would be expected to regress equally.

We can conclude, then, that in at least these two situations remedial reading instruction seems to have helped students learn to read.

If remedial programs can indeed "work," do their results carry over once a remedial student is returned to normal classroom instruction? Although the major goal of most remedial programs is to enable students who are poor readers to progress normally in regular classrooms after a brief period of special assistance in reading, some reading specialists maintain that disabled readers lose many of the skills they have acquired in special programs once they have returned to regular classrooms or have been on vacation.

Five of the fifteen studies under consideration here reported on the achievements of remedial students who had completed their remedial programs and returned to other classes. Two of the five studies (Cashden and Pumfrey, 1969; Pearlman and Pearlman, 1970) were similar in seven characteristics. Both involved children who averaged eight years of age; those children had normal *IQ*s of 90-105; they came from middle socioeconomic backgrounds; their reading levels before remediation averaged 1.4; the teacher-pupil ratios during remediation ranged from 1:1 to 1:4; they were given remedial instruction for more than fifty hours (an average of fifty-four hours in Cashden and Pumfrey, and seventy-six hours in Pearlman and Pearlman); they were taught by certified, experienced diagnostic teachers. Their average reading levels at the end of remediation were 2.6 and 4.0 respectively, with average learning rates during remediation at 1.2 for Cashden and Pumfrey and 3.5 for Pearlman and Pearlman. Two years after their remedial programs, the children maintained learning rates of 1.26 and 1.20 respectively. (These last learning rates were computed from the gains made from

the end of remediation to the end of the following twenty-four months.) These latter rates indicate that the conditions that prevailed in their remedial programs enabled the children to continue to learn to read at slightly better than normal rates after remedial instruction was ended for them. Remedial reading instruction had, apparently, indeed made a difference in their abilities to learn to read.

Profile of successful programs

Which of the many characteristics of remedial programs are crucial? Which can account for such differences? Comparing interventions—techniques, class sizes, materials, program durations, and so forth—might point to some tentative answers, at least.

Lovell, Johnson, and Platt (1962) report programs that were similar to those reported by Cashden and Pumfrey and Pearlman and Pearlman. Although the children in the Lovell, Johnson, and Platt study were slightly older (9.7 rather than 8.5 years, on the average) and slightly higher in average reading level (1.9 rather than 1.4), the critical difference from comparing the three reported programs seems to be length or duration of remedial intervention. The Lovell, Johnson, and Platt program lasted eighteen hours for one group and twenty-seven hours for another—three-quarters of an hour of instruction per week in each case. The duration of remedial instruction in the other two studies was fifty-four and seventy-six hours respectively. Although the learning rates during remediation were similar for the Lovell, Johnson, and Platt and Pearlman and Pearlman studies and the absolute gains in reading were about two years in both, the learning rates twenty-four months after remediation were lower in Lovell, Johnson, and Platt (.55 and .60)—substantially lower than the rates of about 1.20 in the other two studies.

Comparison of two groups in the Cashden and Pumfrey (1969) study confirm this notion that a longer program may yield more permanent effect than a shorter program. One group was given twenty-seven hours, the other fifty-four hours of special instruction. Learning rates during remediation were the same for the two groups, but the post remediation rate for the second group averaged 1.26, for the first group only .73. Apparently, the extended contact (more than fifty hours) with the teacher was important.

It should not be inferred that program effectiveness is therefore directly proportional to program length. Rather, a minimum period of about fifty hours seems necessary, on the basis of comparison among these several studies, to assure long-term gains. There could be many reasons for such a minimum if it indeed exists, but on the basis of these studies, all such reasons would be

conjectural at this point. It might be, for instance, that children in longer programs enjoy more durable benefits because the longer programs allow them more opportunities to apply their newly acquired skills outside the remedial setting. Or, perhaps they learn not only to read, in those programs, but to develop strategies for learning to read, as well. Maybe skill in taking tests—or lack of such skill—offers another explanation for relatively low post remediation rates in programs of less than fifty hours duration. In other words, if the child is given two tests in close succession in a short program, his improvement due to practice on the test might not likely be followed by long-term increases in reading ability.

Balow (1965) concludes that a normal rate of post remediation learning depends on a maintenance program in the post remediation period—continued instruction after the conclusion of the student's remedial program. He showed that post remediation learning rates were .75 for children who received one hour per week supplementary instruction, but were minus .01 for children who did not receive such instruction. However, the reports of Cashden and Pumfrey and of Pearlman and Pearlman did not include any maintenance programs, and post remediation rates in these reports were above normal. One difference between the programs was that Balow's groups received instruction twelve and one-half hours a week for four weeks, while Cashden and Pumfrey and Pearlman and Pearlman's groups received one and one-half and two and one-half hours per week, respectively, for thirty-six weeks. Remedial instruction was distributed over more calendar days in the latter two studies than in the Balow study. That increased distribution may have provided more opportunities for consolidation, for application to outside materials, and for learning to read.

Subnormal post remediation learning rates of minus .15 and of .40 were reported by Heckerl and Sansbury (1968). In this program the children were older (eleven to fourteen), had lower mean IQ scores (WISC Verbal about 80) and lower socioeconomic status than the children that Cashden and Pumfrey and Pearlman and Pearlman studied. The remedial program was also longer, 675 hours, and the teacher-pupil ratio was higher (1:6) than in those other two studies. Consequently, comparison of the Heckerl and Sansbury report does not help us try to define those elements of remedial reading programs that might be said to make a difference.

In summary then, the "successful" reading programs, successful in that they apparently produced normal learning rates in reading that carried over at least twenty-four months after the remedial programs were ended, shared the following characteristics: First, the students in those programs were of elementary age, of normal intelligence, and from middle-socioeconomic backgrounds. The teacher-pupil ratio by which they were instructed in the remedial programs was no higher than 1:4. The teachers were

experienced as teachers or they were graduate students. The programs lasted fifty hours or more. We cannot conclude that all six characteristics must be present before a remedial program can succeed. We cannot even be sure that these factors indeed account for the success of the programs. And, we certainly must not conclude that on the basis of these comparisons, it is futile to expect good results from any program that does not treat elementary students of normal intelligence who come from middle socioeconomic homes, et cetera. Rather, we can say only that these six characteristics were common among those few studies reviewed here as effective remedial programs.

How important is age?

The most appropriate age for remedial intervention is a frequently raised issue. A comparison of two treatment groups in the Pearlman and Pearlman study speaks to the question. One treatment group of 397 children had the same characteristics and received the same basic remedial instruction program as a second group, except that the first group averaged about eleven years of age, with 159 hours of instruction; the second, eight years with 76 hours of instruction. The first's preremediation reading level averaged 2.6 in grade equivalent; the second's, 1.4. Learning rates during remediation averaged 2.70 for the first group that was older, 3.50 for the second group that was younger. Post remediation learning rates (after twenty-four months) were 1.20 and 1.10 respectively. Clearly, both groups learned at normal rates following remediation instruction. Both learned at similar rates during remediation, too, although the younger group was slightly more efficient.

Chronological age at which remediation is most profitably introduced is an issue facing administrators and teachers in all school programs. We compared the learning rates during remediation of programs for primary grades (Pearlman and Pearlman, 1970; Morsink, 1971), intermediate grades (Balow, 1965; Lane, 1974; Strang, 1972; Cashden and Pumfrey, 1969; Pearlman and Pearlman, 1970; Morsink, 1971; Talmadge, Davids, and Laufer, 1963; Lane, Pollack, and Sher, 1972), and post-elementary grades (Marani and Tivvis, 1970, Lane, Pollack, and Sher, 1972). The average learning rates during remediation were 2.6 for primary, 2.8 for intermediate, and 2.2 above sixth grade. Within each age category, the range of learning rates was large, indicating that factors other than chronological age apparently determined the magnitude of the rates. The differences among age groups are probably inconsequential; similarity suggests that the rate of learning to read in remedial programs is not seriously affected by age. However, disabled readers who are relatively old usually have larger discrepancies between actual and expected reading levels than younger disabled

readers. The time required to remediate the older children to a normal level is consequently longer than the time required for younger students. A subtle, but vital, point is that the rate of remediation (or return to normal) is lower for older children than younger children since their learning rates are equal.

Who should teach?

A plausible starting place in the search for critical variables affecting remedial programs is the person providing the instruction. In the assembled group of fifteen studies, nine contained useful information about the teacher and had appropriately homogeneous groups of children. In six studies the programs were conducted by classroom teachers, although they were operating outside normal classroom situations (Steirnagle, 1971; Lane, 1974; Cashden and Pumfrey, 1969; Marani and Tivvis, 1970; Talmadge, Davids, and Laufer, 1963; Pearlman and Pearlman, 1970). The learning rates of children during remediation in these studies were, respectively: 2.4, 1.8, 1.25, 1.0, and 2.6 (experimentals), 3.1 (groups 1 and 2). The average of these is 2.02. With a variety of teacher-pupil ratios, ages of children, methods and initial reading levels reported, it appears that classroom teachers enable poor readers to learn at double the rate for normal children in normal classroom situations. It is interesting to note that the rate of 2.02 is about six times the learning rate of .29, which was the average for these groups prior to remediation.

The effectiveness of graduate student clinicians in a university reading clinic was reported by Balow (1965). The average learning rate during remediation in that program was 5.4. This is higher than the average for programs in which classroom teachers were the primary personnel, but since only one university clinic was represented, the comparison should be postponed until more clinics have reported their programs.

A third source of manpower for reading programs is older students. In two reports (Willis, Morris, and Crowder, 1972; Lane, Pollack, and Sher, 1972) eighth graders, who had been thoroughly trained, tutored young children effectively. The learning rates were 3.0 and 2.4 in these studies respectively, with an average of 2.7, which is higher than the average for programs staffed by classroom teachers. Several cautions should be noted. Very few children participated in both of these programs, ten in the former and eight in the latter. The reading test used in the Willis, Morris, and Crowder study was the Slosson Oral Reading Test, for which the reliability is variable. The stability of the results, and replicability of the findings, are consequently debatable. In addition, the tutors in both studies were intensively trained to implement a preexisting structured program, were supervised, and were encouraged in

frequent group meetings. Finally, the total length of intervention was 37.5 hours in the Willis, Morris, and Crowder study, which is lower than the intervention time of fifty hours that we have earlier suggested seems to be important for producing long term effects.

What conclusions can be drawn from these comparisons? Certainly the published reports do not represent all the programs taught by certified teachers, by graduate students in clinics, or by older pupils. Since less successful programs are less likely to be reported, we may construe these studies as reports of relatively effective projects. At least it is clear that none of these three groups has a monopoly on effective intervention for poor readers. Indeed, briefly, but competently, trained adolescents have tutored with an impact equal to that of university clinicians and classroom and reading teachers. Does this fact suggest that we know enough about reading instruction to transmit teaching skills to relatively uninformed, immature students? Or does it suggest that we know so little about reading remediation that our most highly trained and knowledgeable experts are no better than unsophisticated laymen? The studies imply that effective intervention must be carefully planned by a knowledgeable expert in reading but that the plan often may be implemented by laypersons as efficiently as specialists.

Teacher-pupil ratios

It is widely believed that children with reading difficulties in special remedial programs in small groups need low teacher-pupil ratios to learn effectively; that is, the fewer students per teacher the better, with one teacher to one pupil as the best of all possible arrangements. In attempting to evaluate this belief, the studies were divided into those with low teacher-pupil ratios (1:1, 1:2, 1:3) and those with high ratios (1:4 to 1:8). This dichotomy we justified by the Shaver and Nuhn (1971) study which reported that children randomly assigned to ratios of 1:1 or 1:3 achieved equally well. Other ratios were not included in the Shaver and Nuhn study, so we arbitrarily assigned 1:4 and above to the high category for this review.

From the literature at hand, we drew fifteen groups from eleven of the studies for comparison of the potential effects of high and low ratios (Balow, 1965; Steirnagle, 1971; Lane, 1974; Strang, 1972; Cashden and Pumfrey, 1969; Pearlman and Pearlman, 1970; Heckerl and Sansbury, 1968; Marani and Tivvis, 1970, Morsink, 1971; Kimble and Davison, 1972; Lane, Pollack, and Sher, 1972). In these eleven studies, no group numbered fewer than six (the range was 6 to 808) nor did any remedial program last less than fifty hours.

In the six groups with low ratios, as defined here, the average learning rate during remediation was 3.7; in the nine groups with

high ratios the average learning rate was 1.7. However, the apparent conclusion that low teacher-pupil ratios were, therefore, more effective than high ratios is not really justified, for the IQs of the pupils happened to parallel those ratios rather closely. That is, in eight of the nine groups for which IQ data were available, those groups with low teacher-pupil ratios also had mean IQs higher than 95 (ranging from 99 to 109), and those groups with high teacher-pupil ratios also had mean IQs lower than 95 (ranging from 70 to 93). For groups in which mean IQ was over 95 and teacher-pupil ratio was 1:3 or lower, the average learning rate was 4.4. For groups with lower IQs and higher teacher-pupil ratios, the rate was 1.5. In spite of those facts, we may not infer anything about the influence of teacher-pupil ratios or IQs considered separately. The safest conclusion we can draw from the data at hand is that when a low teacher-pupil ratio was used with children of relatively high intelligence, learning rates were higher than when opposite conditions prevailed.

It should be recognized that the programs with higher teacher-pupil ratios (1:4 to 1:8) for children with lower IQs were not without value. The mean learning rates of children before entering these programs was .26. Since the rate of acquisition during the programs averaged 1.5, it is obvious that the programs accelerated learning rates by a factor of about six, which speaks favorably for the programs. Of course, at the same time it must be pointed out that learning rates in programs with low teacher-pupil ratios and higher IQs were accelerated by a factor of about nine.

Which methods work?

Few words were written by the authors of all these reports on methods of teaching used in the various remedial programs. At a molar level of description, the programs were relatively homogeneous. All included reading instruction using letters, words, sentences and paragraphs. None of them devoted much, if any, time to visual perceptual training, i.e., Frostig materials, or motor coordination training patterned after the Kephart or Delacato models of remediation. The programs were described as "a "programed phonic-linguistic approach" (Lane, Pollack, and Sher, 1972), "lessons comprised of textual material taken from a current graded reading workbook" (Strang, 1972), or "sequential planned materials of great depth to coordinate programing on a visual, auditory, kinesthetic, and tactile basis" (Pearlman and Pearlman, 1970). Other descriptions included "diagnostic teaching," sequential planned materials," "techniques governed by needs of the child, with emphasis on word recognition."

Overall, however, the information from the reports is insufficient for assessing the influence of instructional method as a variable. Tools for recording, gauging, and quantifying teaching

procedures in these remedial programs were generally not reported. Thus, there was no precise measurement of teacher behaviors, pupil activity, or appropriateness of materials.

It is impossible also to compare from these reports the effects of program variables on such different abilities as oral reading, sight vocabulary, and silent reading comprehension. Various tests were used in the various studies: California Achievement Total in three; Wide Range Achievement in two; Gray Oral in two; and Metropolitan Achievement, Nelson Denny, Gates Reading Survey, Schonell Vocabulary, Gates-MacGinitie, Burt-Vernon Word Reading, and Slosson Oral Reading in one each. In still other studies, the names of tests used were not reported. It is possible, however, to use the test results reported in the various studies to compare the results of the studies themselves, since the tests named do intercorrelate at about .5 to .6. Even at that, precision is less than optimal.

There are many shortcomings in this review of the research literature about this one area of instruction—remedial reading. We culled and reviewed only fifteen studies, and few of those are directly comparable with each other in more than one or two characteristics. The review is notably lacking in attempts to gauge such important factors as self-concept, motivation, and atmosphere. Definitions of various program elements too often are vague: Who were the "disadvantaged" pupils in this study, the "middle-class" youngsters in that one? Why were these various groups the ones selected for study in the first place? How much time and interest were invested by parents? Clearly, we are left with far too many questions and too few answers.

But the stated facts offer clues to those seeking to build new remedial programs or repair established ones, clues based on observation. This factor was important according to several studies; it might be important for us. That element proved less than conclusive in that situation; will we need to include it in our planning, our program?

Perhaps the most beneficial result of any such review as this will be the invitation which it extends by implication: to observe as carefully and objectively as possible what we are up to and *why* in our own remedial teaching.

References

Balow, B. "The Long-Term Effect of Remedial Reading Instruction," *Reading Teacher*, 18 (1965), 581-86.

Cashden, A., and P.D. Pumfrey. "Some Effects of the Remedial Teaching of Reading," *Educational Research*, II (1969), 138-42.

Guthrie, J.T. "Reading Comprehension Processes and Instruction," in J.T. Guthrie (Ed.), *Cognition, Curriculum, and Comprehension*. Newark, Delaware: International Reading Association, 1977.

Heckerl, J.R., and R.J. Sansbury. "A Study of Severe Reading Retardation," *Reading Teacher*, 21 (1968), 724-729.
Kimble, R.L., and R.G. Davison. "Reading Improvement for Disadvantaged American Indian Youth," *Journal of Reading*, 15 (1972), 342-346.
Lane, A. "Severe Reading Disability and the Initial Teaching Alphabet," *Journal of Learning Disabilities*, 7 (1974), 479-483.
Lane, P., C. Pollack, and N. Sher. "Remotivation of Disruptive Adolescents," *Journal of Reading*, 15 (1972), 351-354.
Lovell, K., E. Johnson, and D.A. Platt. "Summary of a Study of the Reading Ages of Children Who Had Been Given Remedial Teaching," *British Journal of Educational Psychology*, 32 (1962), 66-71.
Marani, S.D., and J. Tivvis. "A Successful Reading Program for the Disadvantaged," *Reading Teacher*, 24 (1970), 39-41, 65.
Morsink, C.V. "Teaching Early Elementary Children with Reading Disability," *Reading Teacher* 24 (1971), 550-555, 557.
Pearlman, E., and R. Pearlman. "The Effect of Remedial Reading Training in a Private Clinic," *Academic Therapy*, 5 (1970), 298-304.
Shaver, J.P., and D. Nuhn. "The Effectiveness of Tutoring Underachievers in Reading and Writing," *Journal of Educational Research*, 65 (1971), 107-112.
Steirnagle, E. "A Five Year Summary of a Remedial Reading Program," *Reading Teacher*, 24 (1971), 537-542.
Strang, H.R. "An Automated Approach to Remedial Reading," *Psychology in the Schools*, 9 (1972), 433-439.
Talmadge, M., A. Davids, and M.W. Laufer. "A Study of Experimental Methods for Teaching Emotionally Disturbed, Brain Damaged, Retarded Readers, " *Journal of Educational Research*, 56 (1963), 311-316.
Willis, J.W., B. Morris, and J. Crowder. "A Remedial Reading Technique for Disabled Readers that Employs Students As Behavioral Engineers," *Psychology in the Schools*, 9 (1972), 67-70.

In this chapter, the author describes her experiences working with five frustrated, defeated, nonreading third graders. By combining two previously tested methods—simultaneous reading while listening and repeated reading of a familiar text—Chomsky intended to provide these students with needed practice. "I mean to provide printed inputs in large quantity in an interesting and truly accessible form."

This is accomplished through the use of individual storybooks, tape recorders, and teacher-directed activities. Thus, after listening to and reading a story to the point of near memorization, a student is engaged in analytic activities, both written and oral, associated with the text.

The details of this procedure, and the extent to which it was successful in facilitating fluent reading, are described in the following pages in such a way as to make the procedure applicable in similar settings. An important factor to note is the ease with which this method may be used in the classroom.

When you still can't read in third grade: after decoding, what?

Carol Chomsky
Harvard Graduate School of Education

Not long ago a third grade teacher sounded a repeated and familiar distress call. "Can't you *do* something? These kids are at a complete standstill in reading. Nothing I can think of to do seems to make any difference."

She was talking about the five slowest readers in her class at a middle-class suburban school on the outskirts of Boston. The three boys and two girls were eight years old, of normal intelligence (IQs of about 100), with no apparent speech or language problems, but they were all reading one to two years below grade level.[1] In a dynamic school where reading is taught well and for the most part successfully, these children were clearly not being reached.

It was now January of third grade, and their teacher's concern was their persistent lack of progress. Her efforts and those of the reading specialist, which were considerable, seemed to be largely beside the point for them. The same treatment in ever more concentrated doses was getting nowhere, she felt, and destined to continue to get nowhere. As time went by, the gap between these turned off kids and their reading peers continued to widen.

These children were not nonreaders. They had received a great deal of phonics training and had acquired many phonetic

Adapted from Carol Chomsky, "After Decoding: What?" *Language Arts*, 53, 3 (March 1976), 288-296, 314. Reprinted by permission of the National Council of Teachers of English.

Table 1
Children's Grade Scores on Reading Tests Before and After Tape Recorder Work

I. Group test, school administered					II. Individual tests, project related							
	METROPOLITAN ACHIEVEMENT TEST				WRAT	GATES-MCKILLOP DIAGNOSTIC TESTS					DURRELL ANALYSIS OF READING DIFFICULTY	
	Oct. 3rd gr. 1972		Oct. 4th gr. 1973		Reading Subtest	Flash Presentation					Oral Speed	
						words		phrases				
	rdg.	wd. knowl.	rdg.	wd. knowl.	pre	post	pre	post	pre	post	pre	post
Jimmy	2.2	2.2	2.8	2.8	3.0	3.1	3.1	2.9	3.1	4.2	lo/med 2	lo 3
Sue	1.9	2.6	2.3	2.5	3.6	4.1	3.5	3.5	3.0	3.9	hi 1	lo/med 2
Linda	1.9	1.5	2.7	2.7	2.5	3.1	3.5	3.6	5.1	5.4	lo 3	med 3
Rick	1.7	1.2	2.9	2.0	2.6	3.3	2.9	2.9	3.9	3.9	lo/med 2	lo 3
Tommy	-	-	-	-	2.6	3.5	2.9	3.5	3.3	4.2	lo 2	med/hi 2

I. Metropolitan Achievement Test, October 1972 and October 1973. School administered.
II. Individual tests, before and after tape recorder work.
Pretest: March 1973 week one (Grade 3)
Posttest: June 1973 week fifteen (Grade 3)

skills. They met regularly with the remedial reading teacher, with whom they had worked intensively since first grade. A variety of reading tests administered individually in March of third grade, shortly after I started to work with the children, showed average or low-average ability in word recognition and word analysis tasks, and fair to poor reading speed and comprehension. Table 1 gives individual scores on these tests.

After two and one-half years of effort, they could "decode," albeit slowly and painfully. What was so frustrating was their inability to put any of this training to use, their failure to progress to even the beginnings of fluent reading. They hated reading, avoided it whenever possible, and consistently met the many opportunities for meaningful reading, adequately provided for in a lively classroom, with a total lack of response.

All of this occurred in a good school where the methods used worked fine for most children, and where their teachers, the reading specialist, and the school psychologist had invested a good deal of time and energy in trying to help them. It hadn't taken. In spite of their hard-won "decoding skills," the students couldn't so much as read a page of simple material to me. The attempt to do so was almost painful, a word-by-word struggle, long silences, eyes eventually drifting around the room in an attempt to escape the humiliation and frustration of the all too familiar hated situation.

Certainly these children needed something very different from additional phonics or more opportunities to read. In trying to characterize their needs, and the differences between them and

others who do progress to fluent reading, I was reminded of a conversation I had not long ago with a linguist friend. We were talking about the problems that so many children seem to have in learning to read. The conversation turned to our own children's learning to read at home before they entered school. My friend's main contribution to the process, he recalled, was to answer questions. When the children needed to know a word, they would come over and ask. Beyond that, he mostly kept out of the way. I was struck by this because it had been the same with my own children; they each learned to read before entering school, and they did so largely on their own, each one in turn. My main function had been to answer the questions.

Indeed the do-it-yourself ethic was so strong that when I did try on occasion to inject myself into the process, I couldn't get very far. I remember one such incident clearly. My five-year-old son, sitting on the floor trying to read a book, found it a bit of a struggle. He called me over to ask about a word, which I supplied. I remained looking over his shoulder, not wanting to miss a chance to watch and listen to what was going on. Great, I thought, now I'll see how it's done. A golden opportunity. But he lifted his arm and pushed me away. It was a firm shove with a very clear message: Go away and don't bother me. I'm busy.

It seems to me that what is lacking in many children who read poorly in school is just this self-propelled activity. There is far more involved in learning to read than can be "taught" to a child, whether by the linguistic method, phonics, language experience, whole-word, or what have you. The intricate internal organization of the material that has to take place requires active participation on the child's part. It is the contribution of the learner that is crucial rather than particular teaching materials. Unfortunately, these five slow readers had never reached the point where they were caught up in the fascination of learning to read, so that they could go ahead on their own with some confidence. They hadn't managed even to get started.

A child of academic parents, of course, may be highly motivated to participate in learning to read because reading is something that the people around him spend a great deal of time doing. The five children under discussion here, from working class families, had not grown up surrounded by books nor family involvement in the printed word. Their present relation to reading was one of extreme passivity, a reaction perhaps to initial difficulties in learning to read that discouraged them early on. It was as if they had simply retreated. They needed to regain the attitude that reading was accessible to them, that it is an activity worthy of their efforts, and that success is to be expected. In short, they needed to switch from retreat to pursuit.

How to get them started? First, I wanted to assure myself that their phonetic knowledge was adequate. I asked them to spell a number of words, whose conventional spelling they did not know, according to the way they sounded. The children were able to do so; they had no trouble segmenting the continuous stream of speech into phonemic units in the correct sequence. For example, they wrote *human* as *HUMIN*, and *Donald* as *DONELD*. The tester who administered the reading tests in Table 1 reported, "They're not so bad in basic skills, it's mainly that they're so slow. What they need to work on is speed."

The two basic ingredients that had to be provided for these children, it seemed to me, were to *capture their attention* and *make large amounts of material available*. Neither of these factors had operated productively in their background in the natural course of events, so perhaps it would help at this point to introduce them artificially. When I say make the materials available, I don't mean simply to see to it that interesting books were at their disposal. I mean to provide printed inputs in large quantity in an interesting and truly accessible form, so that the mind is engaged in interacting with the print. For a long time these children's minds had been elsewhere when it came to grappling with the written word.

I decided to use *memorization* of a book to get them started in dealing with large amounts of connected discourse. This is what was most seriously lacking in their abilities. Through repeated listening to a tape recorded storybook, while following along in the text, they were to become familiar enough with a book to read it through, either aloud or silently, with complete ease. They were to set their own pace, reading and listening repeatedly to the same story until fluency was achieved. Only after they knew the book thoroughly in this semirote fashion would we begin to do any work on it. At this point, rote recognition would be supplemented with explicit discussion of orthographic features.

Attempts to increase fluency in slow readers by similar methods have been made in the past. The *listening while reading* technique has been used by Heckelman (1969), Hollingsworth (1970), and Neville and Pugh (1974). Their pupils, however, read new material at each listening session rather than concentrating on a particular text. Hoskisson (1974, 1975a, 1975b) has advocated the use of Assisted Reading (reading while listening) as an integral part of normal classroom reading instruction. Dahl and Samuels (1975) have tried repeated oral reading of short passages (without listening assistance) for improving fluency. In the present study, listening and repetition are combined. The pupil listens while reading, repeating a single text until it is perfectly known. In addition, the pupil engages in analytic phonetic work on the familiar material once fluency is achieved.

This memorization of a text is not necessarily as mechanical and empty an activity as it may seem at first. Memorization of books by very young children before they know how to read is a common occurrence among children who are read to a great deal. Young children who watch the page as they are read to often end up knowing a favorite story by heart after repeated listenings. It is then amusing for parents to watch the child sit alone with his book and "read" the story aloud, turning the pages at the proper time, rendering the story accurately and with expression, and looking for all the world as if he were actually reading. Indeed, this type of activity often contributes to early reading, and occasionally turns up as a contributing factor in cases of children who teach themselves to read.

How this discovery of reading feels to the young child is captured in the following personal account by a graduate student who read an early version of this article. She returned the paper to me with a note attached:

> *I* learned to read by memorizing a book—and I just found a copy of the book in a bookstore and had not realized it was as complex (and foolish) as it is. I remember my poor mother being exasperated because she had to read it again and again and again to me. I remember swinging at the playground and repeating the whole story over and over to myself, and I remember it dawning on me which words on the page were the ones that I was saying to myself as my mother read to me, and how exciting the discovery was. When I got to school I used to shoot through the reading out loud sessions....
>
> (Susanna Joannidis, personal communication)

Exciting, yes.

Huey refers to the same type of activity in what he calls the *imitative method* of teaching reading.

> Perhaps we should catalogue still another, the imitative method. In the Orient, children bawl in concert over a book, imitating their fellows or their teacher until they come to know what the page says and to read it for themselves. Many an American child cannot remember when reading began, having by a similar method pored over the books and pictures of nursery jingles and fairy tales that were told to him, until he could read them for himself.[2]

Why not use some of this sort of thing with the third graders who still couldn't read?

These eight-year-olds, in a sense, were starved for inputs. Imagine for a moment a child learning to speak. As he develops language, the infant is surrounded by speech and conversation and is bombarbed in a way by inputs. From the mass of material available

to him from his environment, he selects what he needs to build his language. It is an active process. He operates on the plentiful incoming material by analyzing, organizing, deriving regularities, forming hypotheses and testing them, adjusting hypotheses to accommodate new material,and so on. The child is supremely well equipped to do this. The surprising thing is that he doesn't have to work hard at it. It is easy and natural. If rich and varied inputs in meaningful, two-way communication settings are provided, the child does the rest.

Surely, some of this natural ability that the young child has ought to be available in dealing with learning how to read. The child is by nature able to construct a coherent linguistic system out of a rather haphazard variety of inputs. But in order to use this ability he has to come into contact with large amounts of material. You can't learn to speak and understand without adequate exposure in a meaningful setting and the same would seem to hold true for reading. Children who do learn to read fluently provide themselves with the needed exposure early in the game. Children who cannot read at age eight perhaps could benefit from having the exposure provided artificially.

When it comes to memorizing a book, these eight-year-olds are in a very different position from the toddler who learns a book by heart. These third graders have already had two years of drill in word analysis, long and short vowels, word endings, and blending. They can sound out words and have a fair sight vocabulary. They are beyond needing introduction to the alphabetic nature of the English writing system. What they need is to shift their focus from the individual word to connected discourse and to integrate their somewhat fragmented knowledge. It is the larger picture that they need help with in learning to attend to the semantics and syntax of a written passage and in developing reliance on using contextual clues from the sentence or even longer passages as they read.

The overall plan follows:

1. The children listen individually to tape recorded stories, simultaneously following along in the written text. They read and listen repeatedly to the same story until fluency in oral reading is achieved.
2. The text is then analyzed extensively through a variety of language games, described below, designed to lead the children from rote recognition of the written material toward an active interpretation of alphabetic and phonological features of the writing system. These include sentence analysis, word analysis, and composition involving the textual material.

It will be noted that none of these specific "language arts" types of activities is particularly new. The idea of memorization is an old one, and tape recorded stories are by now commonplace in

many classrooms and libraries. Phonetic drills are everywhere, and free writing is encouraged by many teachers. What is different here, at least in the way we conceived the program, is the particular way all of these separate activities are coordinated into a unified effort on the part of the child. The tape recorder is used to hear a story not just once or twice, but repeatedly, until the story is learned. The child can read in unison with the tape, record his voice, and listen to his own recording at his leisure. The feedback is excellent (see references to Heckelman and Hollingsworth). In addition, phonetic work is carried out on text that is *already familiar*. The memorized story is an unbeatable source of material for individual phonetic work with each child, as will be illustrated below. And finally, the writing of stories, compositions, notebooks, or posters flows quite naturally from knowing these stories well. What is provided here is a means for tying together the various formal language activities of the child so that they lend support to one another and increase their effectiveness.

Procedures

I obtained five tape recorders and two dozen storybooks recorded on tape.[3] The books were a set of children's softcover picture books—some fairy tales and some original stories—attractively printed and illustrated and easy to handle. They ranged from about second to fifth grade reading level. Most of the books were 20 to 30 pages each; a few were considerably longer. The tapes of these stories were lively and interesting recordings of good dramatic presentations, using different voices for the various characters, interspersed with music and song.

I had the children look over the books and make their own selections. I explained to them that they would continue listening to the same book until they knew it well. Since they would work with such concentration on the book of their choice, they were to pick one that was too hard to read straight off but not so hard as to be entirely out of range. Four of them chose among the easier short stories; and one girl chose a long, relatively difficult, book and planned to do just the first chapter (some 20 pages) as a start.

The children were instructed to listen to their books every day, using earphones, following along in the printed text. They were to listen to the whole book through at least once (a 15 minute procedure) and then go back and relisten to any part they cared to prepare more carefully. They could also record themselves reading along with the master tape[4] or record themselves reading aloud independently. I told them that when I came back, I would listen to as much of the book as they had ready.

I came into the school each week to check on their progress and worked with each child individually for about half an hour. The children were also seen individually once a week by a graduate

student working with me as research assistant.[5] In our sessions with the children, we listened to as much of the story as they had prepared, and worked analytically on the passages they knew (see below). We also provided the children with notebooks in which they were to do their writing and drawing connected with the stories.

Progress

The course of their progress was interesting. At first the going was slow. The children took some time to get accustomed to working the machines, and at the start kept losing the place in the book. Although the narration and dialogue on the tape were fairly slow, it took practice for the children to coordinate their eyes and ears and be able to follow in the text exactly where the voice on the tape read them. When the children "got lost," which happened frequently while they were still unfamiliar with the book, they would have to rewind the tape, go back to the beginning, and start over. Once they became familiar with the book, however, keeping the place was no problem. By the end of the first week, this difficulty was over.

It took four of the children approximately twenty listenings to achieve fluency in reading the first book orally. This was about a month's work. They enjoyed it. They were excited by the machines, the new books, and the special treatment, and found the whole process intriguing. One child, Jimmy, was ready sooner. Within two weeks he was able to read his entire book aloud smoothly, with expression, with ease, and with no stumbling. Best of all, his pronunciation was natural—a far cry from the stilted artificial pronunciations that usually characterized his halting attempts to read aloud under pressure and without comprehension.

It was a big event the first time Jimmy read that whole book aloud to me. I was taken by surprise, as he knew I would be. He was so obviously pleased with himself that we went straight to his teacher and he reread the whole thing for her. She could hardly believe it. She called over the reading specialist who had been working steadily with him for almost two years now, and he read the book a third time to all three of us. By this time he was grinning from ear to ear, slightly hoarse and pretty much exhausted. But flushed with success. It was a tired but very buoyed Jimmy who went home that day.

A sample passage from Jimmy's book may be of interest:

There are nine elephants.
Eight of them are clean, but you can see
a very dirty one in the middle.
Who is he?
Yes, he's Gurumpa.
The elephants are having a meeting

in the jungle.
Everybody is holding his trunk
high up in the air
because Gurumpa smells very, very bad.
"Let's give him a bath
and send him to work!"
Now everybody is at the river.
Two elephants are scrubbing
Gurumpa with brushes.
Four elephants are giving him
a shower.
And there is an elephant waiting with a towel.
Everybody is washing Gurumpa.

(The Kindergarten Elephant, Labo Teaching Information Center, Tokyo, Japan, 1969.)

I asked Jimmy how he had managed to learn it so well in so short a time. He shrugged and admitted rather embarrassedly, looking the other way, "I dunno. I liked the book."

Jimmy's teacher had a great deal to say about his progress and involvement. She insisted that I could not possibly realize what this meant for Jimmy, who in his whole life had never read so much as an entire page. The sense of accomplishment for this child, she said, who had known only frustration and failure for two years, was indescribable. Suddenly he could do it! It transformed him in the eyes of his classmates as well. The teacher had him read his book to the class the next day and it was a triumph.

What was the nature of Jimmy's activity as he "read" his book aloud? It was not true reading; he would have been unable to make headway had he not been exposed to the text through repeated listenings—perhaps ten or twelve times. Nor was it pure memorization, for clearly he could not recite the book verbatim without looking at the text. It was a combination of the two. For the time being it would do.

Tommy perhaps put his finger on how it felt to the children themselves. His mother reported that he loved working on the machine. "It helps when I can hear it while I look at it. It makes the hard words as easy as the easy words," she quoted him. And indeed, it did. For Tommy and the others, the struggle and frustration had been taken out of getting started with a book. The child, alone with his book, tape recorder, and earphones, listens and relistens at his own pace until he is comfortable with the book. There is no halting period, no one checking on him, no public humiliation. His difficulties remain a private matter between him and his book. And the activity itself assumes something of the quality of experienced reading. In a sense, the listening permits the child to get the feel of real reading before the ability has been gained. For a turned off eight-year-old accustomed to two years of failure, it is a welcome switch.

The postures these children assumed while listening were interesting. They were comfortable and relaxed, not strained. The children liked to carry their tape recorders to a corner of the room and sit on the floor with them. On occasion, I found one or two of the children in the corridor installed under a big hall table, lying on the floor, listening and reading, oblivious to the clatter of voices and traffic going by. The earphones effectively blocked outside noise and they were in a world of their own. Once I entered the classroom when no one was there; at least the room appeared to be empty. I soon discovered Jimmy stretched out on the couch in the reading corner, listening and reading, engrossed, head bent over his book, looking like a boy lost in a book ought to look. Natural enough for many children, perhaps, but for Jimmy a totally new thing.

In addition to working at school, the children took their tape recorders and several books home for a month or so. They particularly liked having the machines at home and worked with them daily quite faithfully.

In the course of the four month project, all five children made considerable progress. The big hurdle was the first book, which required up to a month's work. Then, the pace picked up and subsequent books took less and less time. By the time the children were on their fourth or fifth book, they were able to finish it in a week's time. By the end of three months, all the children had finished six books or more—six storybooks they could "read" fluently and naturally. Sue, who had chosen the longer, more difficult book at the start, was into her third chapter. She preferred to stay with the one book throughout.

A passage from a two months' progress report on Rick reflects the pickup of pace which is characteristic of all of the children.

> Rick started his first book, *The Snow Daughter*, on January 31. I saw him on February 5, and his attitude was poor, attention wandered, and in general he was fairly negative, although he had selected the book himself from a collection of a dozen or more. By February 8 he was able to read the first two pages slowly, but without expression—it was nothing more than a mechanical performance. At this point, he seemed more willing and better prepared. By February 26 he was able to read the entire book, and he did so, according to my notes, *beautifully*. The expressionless word-by-word stage was apparently over.
>
> On March 5 he took the tape recorder home and started his second book, *The Sky Blue Seed*. By March 12 he could read the whole book. Whereas it had taken him a month to go through his first book, he completed the second in one week. I was astonished at the change of pace.
>
> On March 19 he started his third book, *Jeepta*, and could read all of it by March 22, in less than a week. I started him on his fourth book, *The Ocean Going Orchestra*, on March 26. He still has the

machine at home and I think he should keep it there for a while longer.

In working with him in our weekly sessions, I have noticed, over the two month period, less mind-wandering and a greater ability to take satisfaction in his progress and accomplishments. At the beginning he was so *passive* and resistant—he just hated anything to do with reading and could not bring himself to put his mind to it. Now he participates with some willingness and attention to the task at hand. I find him somewhat less distant in general and am basically encouraged by his progress. Now and then I can even get a smile out of him.

Transfer to new material

What about the children's real reading during these months? Memorized books are one thing, but what transfer is there to new material? Unless the children find themselves better able to undertake reading new material, there has been no real progress. On this score, I received many reports from their parents and from their teacher. All were positive.

The parents' reports of changes in behavior at home were extremely encouraging. Sue's mother was "thrilled with the change in Sue" who was now reading books at home that had been lying untouched for a year or two. Also, Sue was writing stories and had started a cartoon book. She was eager to read to people whenever they would listen. "She even reads to the dog and cat," her mother went on. When she was home sick one Monday, she read three books to her mother. She had even taken to reading the commercials on TV and the cereal boxes on the breakfast table. As her mother summed it up, "She's proud of herself and not ashamed anymore."

As I told Sue's mother at the time, this is the stuff of which reading is made. No one, no matter how competent and well-meaning, could teach her daughter to read in twenty minute sessions at school. It was something she had to take on herself and make a part of her life. Sue was now on her way.

The other mothers reported similar improvement. Rick's mother said he had begun to pick up hockey magazines at home. He even brought a book to the breakfast table one morning. She started to take it away from him (he was already late) but she thought better of it and let him keep it. Linda's mother reported that Linda was picking up more books at home. Three different mothers repeated the story about reading the cereal boxes on the breakfast table.

And at school the same increase of interest was displayed. The children picked up books and sat down to read them. Their teacher emphasized the changed behavior. The children went so far as to choose reading during *free time*, she reported, which was entirely new for them. Sometimes they read to each other and sometimes they read silently. They wrote and illustrated stories and

hung them on the display board. Altogether the change was dramatic.

There were several occasions on which the teacher said to me that even if only *this* had come out of the whole project, the entire thing would still have been worth it. One such time involved Linda and an original story she wrote based on the book she was listening to on tape. She brought her story to her teacher. "Read it to me," asked the teacher. Linda's response was "Oh, no! I couldn't!" After much urging, she finally agreed and did read her story out loud. She was sure it was "terrible," and it took a lot of praise to convince her otherwise. The school psychologist happened by and the teacher asked Linda to read it to him, too, which she did. Between them, they tried to get her to enjoy the fact that she had done a really fine piece of work and to take some pride in it. The crowning moment came when Linda finally handed her story to the teacher with a smile of anticipation "Now YOU," she announced with something of a flourish, "read it to ME."

Linda copied that story onto large construction paper, illustrated it, mounted it, and hung it on the display board. The next day she brought her mother to school to see it.

Linda wasn't concerned with "correct" spelling and wrote quite phonetically much of the time. Since these were her first attempts at original writing, we didn't insist on correct spelling, only that her spellings reflect the way she actually pronounced the words. Thus when she wrote SEEDY for *seed* I had her go back and figure out that the Y didn't belong, but when she wrote TIURD for *tired* I left it alone. This seemed like a reasonable compromise, temporarily. Sue, on the other hand, was interested in spelling everything the right way. She asked to be corrected and her final product was always letter perfect.

The writing that the children did seemed to play an important role in their progress. They wrote stories, question and answer sequences, sentences using words from the stories, and some even made up little tests for themselves. Whether the writing itself contributed to the progress in reading, or whether the greater fluency in reading made writing more accessible to the children, is hard to tell. At any rate, it became apparent along the way that the two activities were closely interrelated, in that the children who progressed the most progressed in *both* reading and writing.

There was an interesting sidelight that appeared after Rick had had his tape recorder at home for several weeks. His younger brother, a second grader, used the machine during this time and went through several of the story books. His teacher, unaware of the circumstances, noticed a spurt in his reading at school and commented to the mother that "Scotty seems to be doing much better in reading all of a sudden." The mother told me about this with some amusement and some wonder. I checked with the teacher and, yes,

he had noticed a rather sudden improvement and hadn't known what to attribute it to until he talked with the mother. Apparently Scotty, on his own, had derived a benefit from the listening and it showed up noticeably, and unexpectedly, at school.

Shift in attitude

Just what accounts for the shift in attitude that these children displayed? Surely there was added motivation when someone came in from the outside with shiny machines and new books and made the five children the focus of special attention. The other children in the class kept asking, "Why can't we use the machines? Why do they get to use them and not us?" For once, the five had something good that the others wanted. But there was, I think, much more involved than just a flurry of newness and special treatment. The teacher and school psychologist felt strongly that the built-in feature of success was most important. There could be no failure in undertaking a book—you simply listened until you knew it. Your difficulties were surmounted early and privately, and *you could not fail.* For these children, who had become so accustomed to failure, the relief was staggering. Finally, they had something they could *do*, instead of the many separate, sequenced reading tasks that in the past had been sources of frustration.

Kagan states it very well:

> . . . individuals will cease investing effort in a problem if they doubt their ability to solve it—if they have no expectation of success. Much of the time the only way a child can tell if he is progressing adequately toward a goal is by checking to see how other children his age are doing. If he is advancing at the same rate as they, he feels confident and continues to work. If he perceives that he is far behind, he is apt to conclude that he is incompetent and cease investing effort.

This would seem to be a good description of what had happened to our five children. Having fallen behind, they remained there until a form of success made it worthwhile once again to begin investing effort.

Jimmy expressed his new-found confidence at a weekly conference on reading his teacher was holding with the children. On the Wednesday morning in question, some two months after starting the project, Jimmy couldn't wait. "When ya gonna conference me?" he started in. His teacher covered up her surprise and said she planned to do it Friday. Evidently, that wasn't soon enough for Jimmy, and the teacher agreed to see him after lunch.

Improvement in learning to read

In addition to added motivation and guaranteed success, what in the nature of the activity itself contributed to improvement

in learning to read? The children obviously began to find reading easier and more pleasurable because they sought out other books and began to interact with print in the world around them (cereal boxes and TV commercials). I think the main factors are the radically increased inputs and concentrated attentiveness. The *child is bathed in inputs with which he interacts.* The mind is active, not passive, and the material on which it can go to work is available in large amounts.

Support activities

The overall purpose of the weekly work sessions was to lead the children from rote recognition of the written material toward an active interpretation of features of the writing system. The activities were in the nature of games involving sentence analysis, word analysis and synthesis and writing involving the textual material. At the start we considered these supplementary activities very important as a follow up to the children's sessions with the tape recorder. We viewed the listening as preparatory to the work that they would then do on the material that they had "memorized." The listening took care of the drill; our work would make the material functional.

The goal was to move from sight recognition of whole sentences to exploitation of the alphabetic and phonological nature of the orthography. To some extent, the memorization allowed the children to treat English as if it were Chinese, as if they were interpreting an ideographic writing system rather than an alphabetic one. They needed to develop their sense of the relations between letters and sounds and to build up a system for exploiting the whole complex network of these relations in dealing with new material.

Let me stress that when we met with the children in our weekly sessions we worked only on pages that they could already "read" fluently in their semirote fashion. One type of activity was identifying words that the children could not recognize out of context. I isolated words on the page by forming a small window in the center of four index cards. This makeshift frame covered everything on the page except the word or phrase in question, and could be moved about easily with the window size adjusted as needed. We moved about on the page in this fashion, exposing different words until we hit one that the child couldn't recognize. Removing the index cards allowed the child to scan from the beginning of the sentence which of course gave him the word. He then copied the word onto a flash card. In this fashion the child built up a personalized inventory of flash cards for each book, a page yielding anywhere from no flash cards up to seven or eight.

The collection of flash cards that each child gradually accumulated was then available for his own specific purposes: word recognition, dictionary construction, composition work, etc. What-

ever games one wishes to devise with the flash cards serve the purpose for word recognition, it seems to me. Sometimes the child would simply flip through the pack and read the words. When he missed a word he would find it on its page in the book in order to identify it. It was always fun to see the light dawn as he located the word in context on the page. A surprised "Oh!" was usually the reaction when he realized he knew it after all.

Sometimes the children enjoyed speed games with the flash cards. They could record themselves reading through the cards rapidly, and try to race against themselves on a second round. Since they could hear the old recording while making a new one, they could keep this up until top speed was reached. The tempo of this game was always lively.

Sometimes we worked with the page as a whole. I would say a word and the child would find it on the page. Rick at one point suggested that *he* read the word and *I* find it. It takes some doing, I discovered, to find a word on a page of some ten lines, and it certainly feels like a useful and relevant activity. Or we would choose one line and I would give him clues for finding a specific word: "Find the word that rhymes with *small*" or "Find the word that begins (or ends) the same as *sprinkled*." Or again, he would create the tasks for me to perform. Any games of this sort that focus attention on individual words or parts of words when the context is familiar are useful.

We also spent a great deal of time composing and analyzing words using a set of wooden letters. When a child had trouble with a specific word, I would have him build it up on the table with the letters. Then we would work on why it said what it did. For example, Jimmy once read the isolated word *lonely* as *lying*, which appeared elsewhere on the page. I had him compose *lonely* and *lying* with the letters. Many activities are possible here. For example, take LONELY. First, he broke it into syllables. Working with LONE, I said "What do you have to do to it to change it to *bone*? No problem. He replaced the L with a B. He was able to change it to BONE, CONE, TONE, and then to LINE, LANE, etc. This can be continued as long as it is interesting. LYING was easy. He knew the *-ing* ending. It was easy for him to change it to TRYING, SPYING, CRYING, etc.

These two words, LONELY and LYING, happened to present a convenient opportunity for bringing out the importance of context for interpreting and pronouncing certain letter combinations. There on the table were the two words which Jimmy had separated into syllables.

LONE LY

LY ING

The two morphemes *-ly* and *ly-* happen to be spelled alike, which probably contributed to Jimmy's error. However, the *ly* of course represents different forms in these two words. We could discuss the

difference in its function in the two words, how you tell the two forms apart according to context, and the resulting difference in pronunciation. This is useful because it brings out the reader's need to know what the word is before he can decide how to pronounce its component parts. I think this is an important point. It places the emphasis on the word without forfeiting reliance on alphabetic and phonological principles. I think this emphasis is important in general for children learning to read and in particular for these children who had been so heavily over-drilled in the particular brand of phonics that produced, time and time again, such artificial and mechanically mouthed sequences as buh-ee-tuh for *beet* and uhll-eh-tuh for *let*.

The spelling with the wooden letters was popular and we used it a great deal. Tommy, for example, kept reading *even* for *never*. I had him compose NEVER, and then show me how he would change it to EVEN. He changed it back and forth several times for practice. Then he closed his eyes while I set up one or the other of the words, opened his eyes on signal, and yelled out the word as fast as he could. I tired of this game well before he did.

The children came by and played with the letters on their own, as well. One day, I noticed Jimmy and a friend surreptitiously doing something at the window near the letter box. When they ran off giggling I discovered they had spelled out AFTER SCHOOL HOCKEY on the window sill. (Their favorite activity, I learned later.) Since there were only two alphabets in the set, they used up the O's in writing *school* and had none left for *hockey*. They had improvised by using the two J's end to end to form an elongated O: H∫∫CKEY. Quite innovative.

As the months went by, we began to put less and less stake in the importance of the follow-up analytic work. After a time the children seemed to be moving along all by themselves. By the third month or so, the most they wanted to do when we met was to read the pages that they had ready, talk about the story, show what they had written, do some more writing perhaps, and move on. It began to appear superfluous and somewhat arbitrary to put them through a preplanned analytical type of work session that they almost did not need. My research assistant independently reported the same shift in her assessment of the children's needs. This change appeared to be a sign of progress and we found it encouraging. So the sessions after a while became times when the children simply read their books aloud and discussed the story, did some writing, or discussed stories that they had written. Writing became the dominant theme of the sessions for several of the children, and they produced quite a bit of good writing over several months' time. Occasionally, we also read to each other from new books that I would bring along for them to take a look at.

Conclusion

How then to assess the overall impact of this trial approach to improvement in reading? All in all, I felt that the children had been given an access to reading that they had not managed to provide for themselves earlier in the game. Mechanical as the idea of memorization may seem in itself, it gave these children practice in reading connected discourse, and put them in touch with a variety of books. They had a feeling of success right from the start, and a sense of progress as book after book was added to their repertoire. No doubt, the fact that the tape recorder work was supplemented by frequent independent writing and language games (analytic phonetic work) based on the familiar texts helped to keep the reading from deteriorating into a mechanical activity. Throughout, the repeated listening was treated as a background preparatory activity which served the function of getting the child a step closer to the real activities that contribute to literacy: interacting with print in the world around him, willingness to undertake reading new material, and original writing.

During the four months the children were involved in this project, their passivity about reading declined dramatically, confidence increased, and they began to pick up new books of their own choosing. Fundamentally, the basis for progress now existed in the form of confidence to move forward on their own. And this, to my mind, was what these children needed most: a chance to get started—the chance they had missed the first time around.

Footnotes

1. On the Metropolitan Achievement Test administered in October of third grade, their grade-level scores ranged from 1.7 to 2.2 on the Reading subtest, and from 1.2 to 2.6 on the Word Knowledge subtest. See Table 1 for individual scores.

2. I am indebted to Helen Popp for bringing this passage to my attention.

3. I am indebted to the TEC Company, Ltd. (Tokyo English Center) of Japan, who provided from their Labo materials the tape recorders, tapes, and books for this project. These materials are designed for use in teaching English as a foreign language to Japanese children, but many of the tapes and books are quite suitable for children who are native speakers of English.

4. Similar to the neurological impress method of teaching reading reported by Heckelman (1969) and Hollingsworth (1970).

5. I am grateful to Betsy Madsen for long hours of work with the children, and for her imaginative ways with them, especially in the area of independent writing.

6. Pre- and posttest scores on several reading diagnostic tests (See Table 1) at week one and week fifteen of the study showed some encouraging gains. On the Wide Range Achievement Test (WRAT) Reading subtest, the children averaged a gain of five months; the Durrell Analysis of Reading Difficulty showed gains in oral speed of several months to one year. Interestingly, the Gates-McKillop subtest *Phrases: Flash*

Presentation (alternate form given on posttest) showed the highest gains of all—up .9 in grade score for two children, and 1.1 for a third child (cf. Table 1). On the Metropolitan Achievement Test administered in October of fourth grade, five months after the end of this project, the children showed grade-score gains of from .6 to 1.2 over their scores of a year earlier (cf Table 1). While these fourth grade scores are still well below grade level, half of them show a substantial increase in rate of progress during third grade, as compared with grades one and two.

References and Notes

Chomsky, Carol. "After Decoding: What?" *Language Arts*, 53, 3 (March 1976), 288-296, 314.

Dahl, Patricia R., and S. J. Samuels. "Teaching High Speed Word Recognition and Comprehension Skills," unpublished manuscript, 1975.

Heckelman. R.G. "A Neurological Impress Method of Reading Instruction," *Academic Therapy*, 4, 4 (1969), 277-282.

Hollingsworth, Paul M. "An Experiment with the Impress Method of Teaching Reading," *Reading Teacher*, 24, 2 (November 1970), 112-114.

Hoskisson, Kenneth. "Should Parents Teach Their Children to Read?" *Elementary English*, 51, 2 (February 1974).

Hoskisson, Kenneth. "The Many Facets of Assisted Reading," *Elementary English*, March 1975a.

Hoskisson, Kenneth. "Successive Approximation and Beginning Reading," *Elementary School Journal*, 75, 7 (April 1975b).

Huey, Edmund Burke. *The Psychology and Pedagogy of Reading*. New York: Macmillan, 1908, 274.

Kagan, Jerome. "The Poor Are Educable," *New York Times* Annual Education Review, January 16, 1974, 81.

Neville, Mary H., and A.K. Pugh. "An Empirical Study of the Reading While Listening Method," paper presented to the Eleventh Annual Conference of the United Kingdom Reading Association, 1974.

The message offered the classroom teacher in the following chapter reflects a shift in emphasis in reading research. This shift, from attention to the special class back to the regular class, follows in the wake of notions associated with the term *learning disabilities*. First noting the arrival, approximately ten years ago, of the learning disability specialists, Jeanne Chall proceeds to document the eventual prominence of such learning disability ideas as the neurological, the physiological, and the psychoneurological as causal factors in learning difficulties. These learning specialists subsequently devised tests and treatments based on visual-perceptual, visual-motor, and auditory-perceptual factors.

Has this practice of viewing reading problems as a symptom of a deficiency in one of the above factors, and then proceeding to provide training in the identified area, resulted in the improvement of reading skills? Has any method produced such an improvement? Are there effective means of diagnosis, prevention, and remediation of reading difficulties?

In the following essay, Dr. Chall discusses all of these questions. She suggests that, where teaching to the "deficient" visual, auditory, and perceptual motor skills has generally failed, treatment based on teaching reading has succeeded, and asks us to take another look at the regular teaching of basic reading skills.

A decade of research on reading and learning disabilities

JEANNE S. CHALL
Harvard University

About ten years ago, a change began to take place in the predominant way in which children with severe reading disability were viewed. The change started with theories of causation and soon spread to preferred methods of diagnosis and treatment. It then spread to terminology and to the preferred professional preparation of those working with such children (Chall, 1966).

At the risk of oversimplifying, let me identify these changes in direct terms as a means of introducing the issues. The needed qualifications will follow.

Ten years ago when a child had great difficulty with reading and had normal or higher mental ability, if the classroom teacher wanted professional help, she usually referred the pupil to a reading clinician, to a psychologist, or to both. If the problem was very severe, the teacher probably referred the child to other specialists as well in order to get at the "cause(s)." The assumption was that usually more than one cause is involved (Helen Robinson, 1946; Roswell and Natchez, 1971; Weiner and Cromer, 1967).

The usual recommended treatment was to alleviate, where possible, the conditions that might be causing the reading problem: visual problems, emotional difficulties, etc. In most instances,

remedial reading instruction was also recommended and, where the budget permitted it, on a one-to-one or small-group basis. Diagnosis and treatment were performed by a reading specialist and/or a clinical-school psychologist who might also have a specialization in reading.

Although a thorough diagnosis would involve many professionals (pediatrician, opthalmologist, speech and hearing specialist, psychiatrist), the major collaboration was between two professional groups—clinical-school psychologists and reading specialists.

About ten years ago, a collaboration of two other professional groups began to take on prominence—neurologists and special educators. Their causal theories of reading disability tended toward the neurological and physiological. There was less emphasis than before on emotional and educational factors. Further, they sought for those "basic" underlying psychoneurological factors that seemed to impede learning to read. They also preferred new labels, eschewing the term *reading disability* for such terms as *specific language disability*, *dyslexia*, and the more comprehensive term, *learning disability*.[1]

The underlying factors first reported in the literature on learning disability were the visual-perceptual and visual-motor factors. Somewhat later came the concern for the auditory-perceptual factors. More recently, concern is being placed on patterning, sequencing, and integrating auditory and visual factors (Hurwitz, et al., 1972; Jansky and de Hirsch, 1972).

The tests used to diagnose children with learning disability increasingly included tests designed to measure these visual and auditory factors: the ITPA, the Frostig, and various tests of auditory perception and blending. Psychological tests to estimate the degree of emotional difficulties were generally dropped although in the 1940s and 1950s, they were considered an essential part of the test battery.

With regard to remedial treatment of children with a reading/learning disability, two major schools of thought appeared about ten years ago. One has tended to favor a remedial (or preventive program) based on strengthening the deficit visual-perceptual and/or auditory-perceptual abilities (Silver and Hagen, 1967). Related to this school of thought are those who urge that the perceptual training be done in the context of the academic area at issue (Johnson and Myklehurst, 1967; Kirk and Kirk, 1971).

The other school of thought recommends "teaching to the strength" rather than to the deficit or weakness. Thus, it was recommended that those children stronger in visual than in auditory factors be taught to read by a sight approach. Those stronger in auditory than visual should be taught by phonics. Those

weak in both should use a kinesthetic approach (Johnson and Myklehurst, 1967; de Hirsch et al., 1966). Recent studies attempting to validate this particular view have been disappointing generally (Park in press).

What has happened to these "new" views during the past decade? Has the research confirmed them? These are big questions and I will attempt to answer them in the remainder of this chapter by breaking them into more specific ones.

1. Who is doing the research in reading and learning disabilities? What is being researched? Have there been any changes over the past ten years?
2. Has there been a change in the past ten years in the concept of reading and learning disability? In diagnosis? In prevention? In treatment?
3. Is there a change in the theories of causation?
4. What stage of reading development seems to be of greatest concern to researchers and to practitioners?

My immediate data come from recent issues of *The Journal of Learning Disabilities* and, to a lesser extent, *The Reading Teacher.* Each is the major publication of the two principal associations concerned with learning disabilities and reading—the Association for Children with Learning Disabilities and the International Reading Association.

I relied, also, on my own research and clinical experience, and particularly on my experience during the past ten years on national and state advisory committees concerned with reading and reading disability (e.g., National Advisory Committee on Dyslexia and Related Reading Disorders, Reading Committee of the National Academy of Education, and the National Reading Council—the private sector committee of the Right to Read Effort).

Who is doing the research? What is being researched?

In a count of the references under "Learning Disorders" and "Reading Deficiency" in *Psychological Abstracts* for the years 1962 to 1972, Black (1974) found a twelvefold increase in both over the ten-year period (from a total of 28 articles in 1962 to 337 in 1972). Of even greater interest is the increase in the Learning Disorder category. In 1962, none of the 28 articles was classified under the rubric of Learning Disorders. But by 1970, most were so classified—(113 out of a total of 204). In 1972, the Learning Disorders rubric still leads with 174 articles compared with 163 under Reading Deficiency.

Since learning disability specialists tend to prefer the term learning disability or learning disorders (compared to the reading specialists who prefer the term reading disability or deficiency), it may well be that this reflects who has been doing the research and

within what framework. It may very well be that in less than ten years, the learning disability group has come to a point of equal, if not greater, prominence when compared to the reading disability group.

Black (1974) made a further classification of the content of all the articles, those classified as learning disorders and as reading deficiencies. He classified them as to whether they were concerned primarily with description and diagnosis, with etiology, or with remedial procedures. He found the great bulk concerned with description (of testing, diagnosing, correlating test results), less than 10 percent were concerned with etiology, and about 25 percent with remediation. Further, there was no appreciable change in the content emphases over the ten-year period.

To sum up, who is doing the research in reading and learning disabilities? Both reading and learning disability specialists. It seems, however, that the learning disability specialists, who entered the area more recently, are probably the greater contributors.

What are they researching? They are describing, testing, correlating, and predicting reading and learning disabilities. Only a fraction are studying what the best treatments are for such children. A still smaller fraction are researching the causes of reading and learning disabilities.

Has there been a change in the concept of reading and learning disability? In diagnosis? In prevention? In treatment?

It seems that there has been a decided shift, primarily in the cause and the symptoms of the disability. More specifically, the shift has occurred in the faith that is placed in the basic psychological processes as the underlying cause, as the focus for disagnosis, and as the way to remediate reading and other learning disabilities. Ten years ago diagnosing and treating the basic psychological processes, such as visual and auditory perceptual abilities, was the new hope (Chall, 1966). As will be seen, a deep disillusionment seems to have set in among learning disability researchers and clinicians.

The disillusionment began as early as 1971 with the Masland and Cratty position paper for the National Advisory Committee on Dyslexia and Related Reading Disorders, published in the *Reading Forum*. They concluded from an extensive evaluation of the research on the association between nonacademic factors (visual perception, visual-motor coordination, etc.) and reading achievement that the association had not been sufficiently demonstrated. With regard to training these nonacademic factors in order to bring about improvement in reading and related skills, they concluded that the evidence was inconclusive. Indeed, most studies that compared reading instruction with perceptual training as a means of remediation found reading instruction more effective. Thus, in

the terminology of the learning disability field, the research—such as it was—indicated that direct training of reading and spelling (which some referred to as "symptoms" of the learning disability) achieved better results than training the "underlying psychological processes."

Essentially, the same conclusion was reached by Barbara Bateman in her position paper in the same volume. She recommended strongly that the extensive diagnoses as well as the treatment favored by those who held to the "underlying psychological process" theory be replaced by task analysis and direct teaching of skills (Bateman, 1971). Her position strengthened even further in a recent publication (Bateman, in press).

In a review of research on the relation between measurable perceptual-motor skills in young children and academic success, Saphier (1973) concluded that perceptual-motor testing and training are still at a "primitive state of the art." They are one of several vital areas of competence for a child but "by no means as preeminent as the mass of literature devoted to it suggests."

Black (1974) questions the very essence of what is the "cause" and what is the "symptom." He reports the results of an empirical study of eight-year-old learning-disabled children, whose test performances classified them as high and low perceivers.

He found no significant correlation in his sample of eight-year-olds between visual perception, as measured by the Frostig, and reading and spelling, as measured by the WRAT (Wide Range Achievement Test). After considering the possibility that at age eight visual perception may not be as important for reading as general intelligence, he stated:

> In light of these results, the heavy emphasis upon visual-perceptual evaluaton and remediation with learning disabled children does not appear to be the efficient road to academic success. As the inferior performances of the high perceivers on both the WRAT spelling and reading subtests indicate, the current policy of offering remedial programs to children based upon the presence of visual-perceptual dysfunction, as measured by the Frostig test, appears untenable. ... The placement of children with learning disabilities in programs for remediation should be based upon the presence and nature of the achievement problem, not upon the presence or absence of a concomitant factor such as a visual-perceptual dysfunction .... Once placed in remedial programs, the emphasis for these children should be upon good teaching of reading and related academic skills, rather than on remediation of concomitant splinter skills. *The most efficient way to remediate reading problems would seem to be the teaching of reading* (italics mine. p. 182).

We seem to have here a complete reversal of what is to be considered the basic factor, and what the splinter skill- -in less than ten years.

Tanis Bryan (1974) questions the very concept of dyslexia and learning disability in a review of the empirical research comparing learning disabled children with academically successful ones. She finds little data to support most of the characteristics used in the literature on learning disabilities. Among the characteristics that did *not* discriminate in the research were: simple perception and discrimination, hyperactivity, and neurological deficits.

Behaviors that *did* discriminate were: ability to pay attention, difficulty with complex tasks that make heavy demands on language skills, and information organization.

A similar questioning has taken place with regard to auditory factors. Hammill and Larsen (1974), in a review of studies using correlational procedures to study the relationship of selected auditory perceptual skills and reading ability, failed to validate "that particular auditory skills, as *measured*, are essential to the reading process, and that *many* children fail to read proficiently because of auditory perceptual deficits" (p. 434).

If there is no demonstration from the existing research that visual-perceptual, visual-motor, and auditory-perceptual training are effective for improving reading achievement, what is effective? Increasingly, the articles in the *Journal of Learning Disabilities* say it is reading—teaching reading.

There were many signs of interest in the teaching of reading, and in remedial reading procedures for children with reading and learning disability. Thus, Richardson, Winsberg, and Bealer (1973) present the results of an assessment of two methods of teaching phonics. Both were found effective in a state hospital with eighteen neuropsychiatrically impaired children. The one developed by the authors was found to be more effective, as is usually the case in the reading literature as well.

Silberberg, Iversen, and Goins (1973) present a report of a comparison of four methods of reading remediation: an auditory-phonic, a visual, a multimodality Orton-Gillingham, and a kinesthetic.

For children at the third grade level, the auditory-phonic method seemed to work better than the kinesthetic, visual, or multimodality Orton-Gillingham. However, it is interesting to note that the authors reject this finding.[2]

In the learning disability literature, one of the few references to the rich knowledge base that exists in the field of reading is Richardson and Bradley's article (1974) in which they present their procedure for learning-disabled children called ISM; A Teacher Oriented Method of Reading Instruction for the Child Oriented Teacher. I quote from them:

> The principles and techniques outlined here represent no advanced scientific breakthroughs in the fields of education and reading. The

ISM is simply a solid approach to reading instruction based on the ideas and research of numerous experts such as Gates, Chall, Bateman, Montessori, and Ashton-Warner. The unique aspect of the ISM is its fundamental premise that teachers themselves are the key to instruction and, given sound principles of education, are fully capable of teaching children as individuals" (Richardson and Bradley, 1974, p. 351).

As a summary to the growing shift away from referring to a reading problem as a symptom, let us turn to Bannatyne's review of *Power in Perception for Young Children: A Comprehensive Program for the Development of Prereading Visual Perceptual Skills.* He writes:

> This is a useful source book of ideas and lessons. Of course, the training the children receive is much wider than just the "visual perception" mentioned in the title of the book. But I hasten to add that no one has yet shown that children on a program such as that described above will learn to read more quickly than those on other types of programs.
>
> In fact the evidence is that programs which teach language structure, phonetic skills and phoneme and articuleme sequencing *directly*, such as the Bereiter and Engleman program reviewed earlier, are more successful in laying the necessary foundation and in building academic success. Consequently, even such a broad and comprehensive program as "Power in Perception for the Young Child" should not be expected to prepare the preschool or kindergarten child specifically to *read* (Bannatyne, 1974, p. 337).

Is there a change in the causes attributed to reading and learning disability?

Here, too, I think there are changes. Generally, in the 1960s there was a tendency to define severe reading disability, or learning disability, as a characteristic stemming not primarily from environmental deficiencies but from neurophysiological and neuropsychological factors. Thus, by definition, a child with a reading problem, of normal or even superior intelligence whose home background was not sufficiently stimulating and/or whose schooling had been inadequate, would not be classified (theoretically) as having a learning disability. By the definition widely held in the 1960s, a deficiency in an underlying psychoneurological factor would be a legitimate reason for classifying a child as having a learning disability.

If this interpretation of the definition is followed strictly, it is possible that children of lower socioeconomic status, bilingual children, and children who go to very poor schools would be eliminated from consideration for special learning disability programs. Indeed, it has been observed, as well as surveyed, that

dyslexia and learning disability may be middle-class diseases (Hurwitz et al., 1972).

The learning disabilities literature of the past decade tends to reflect the emphasis on the individual and the neurophysiological factors. Thus, in their first predictive study, de Hirsch and Jansky said little about the environmental conditions that might have lowered or raised the probability of reading failure among their high risk children (de Hirsch et al., 1966). In their replication study of 1972, Jansky and de Hirsch associated the quality of teaching that their high risk children received in the first and second grades with their achievement. They found a very strong interaction between the two. Thus, Jansky wrote:

> In classes taught by adequate teachers (as rated by principals) 23 percent of the children failed. When the teachers were characterized as poor, the failure rate rose to 49 percent. Teacher competence seemed *to make more difference for the weaker group*—the white boys, the black children, and the Puerto Ricans—than for the white girls (Jansky et al., 1972, p. 83).

Thus it seems that one can have a learning disability or a reading disability with one teacher and not with another. The number of learning disabilities can be cut *in half*, if classroom teachers are competent—even though the children are lacking in "basic psychological processes," as determined by the extensive test battery.

The reading and learning disabilities research published during the past ten years indicates that with regard to diagnosis and treatment, there seems to be a return to interest in the academic skill itself, with a lessening of interest in underlying psychological factors. There seems to be one exception. Theories of causation of reading and learning disability seem to be influenced more and more by the neurosciences (Geschwind, 1974; Gardner, 1975; Chall and Mirsky, 1978).

What stage of reading development seems to be of greatest concern?

Most, if not all, of the articles on reading and remedial reading in the *Journal of Learning Disabilities* are concerned with beginning reading at first through about third grade level. The remedial reading programs reported may be for older children, even adolescents or young adults, but the programs are on a primary reading level. Further, within this limited range, the emphasis seems to be primarily on decoding in beginning reading.

If the ultimate goal of remediation is helping children function on a level commensurate with their mental abilities, is it

possible that their problems are "cured" at the decoding stage? Are there no problems in their reading comprehension? Are there no problems with reading at the intermediate levels? At advanced levels? In their use of reading for study and for recreation and inspiration? If we were to study them longitudinally, we might well find that problems with early reading and learning disabilities continue (in different forms) in upper elementary, high school, and even college.

It would seem that we should be gaining insights into these problems, and into the difficulties experienced in the more advanced stages of reading. This is necessary if we are to understand how to help the millions of students in high school and college who are having extreme difficulty keeping up with their assignments, and who are falling short in achieving on a level equal to their intellectual potential.

The Reading Teacher carried relatively few articles, within the same period of time, on the topics of reading and/or learning disabilities. Indeed, this would be the major observation: The main journals of the two largest and most influential organizations concerned with reading disability contain a disproportionate number of articles on the topics of reading and learning disabilities, with the *Journal of Learning Disabilities* carrying the large bulk of them. *The Reading Teacher* had relatively few articles on the topic, but those found tended to show a deep concern for issues that have surfaced during the past ten years. Thus, in the article by Allington (1975), great concern is expressed about labeling children. If taken to its ultimate conclusion, the overall position would lead to a situation where children would not be classified, graded, or assigned a description, since (according to the author) these become rigid and lead to destructive treatment of those labeled.

Another paper by Niensted (1975) analyzed the plight of the remedial reading teacher in a school system with a learning disability orientation. The differences in language, concepts, and assumptions are presented with charm, humor, and style. It ends with a poignant plea for understanding and cooperation between those from the reading emphasis and those with the learning disability emphasis.

What does the recent research on reading and learning disability say to the classroom teacher?

The main lesson, it seems to me, is that the teacher is of tremendous importance in preventing and in treating children's reading and learning disabilities. When so much of the writing during the 1960s concerned the profound deficits of children with learning disability, many classroom teachers felt there was little for

them to do to help the child. In the zeal to find the "new," the "true," the "real" causes and to treat them, the constructive role that a classroom teacher could play was too often overlooked.

The research of the past ten years suggests, instead, that good teaching is probably the best way to help the child. In fact, it seems that direct training in the academic skill (reading, spelling, writing) is, in most instances, more effective than perceptual training for the underlying psychological factors such as the visual and auditory perceptual. Thus, the classroom teacher becomes a focal point in the treatment. Even if the child is fortunate enough to receive special individual help from a reading or learning disabilities specialist, the classroom teacher plays a central role in prevention and in treatment through collaboration with the remedial or resource teacher.

The teacher is usually the key person to recognize that a problem exists. Since there is ample evidence that the later the child is diagnosed and treated the more difficult it is for him "to catch up," it would be essential that the classroom teacher has the skills for identifying such children.

The teacher can take pride in the importance of the quality of his/her teaching not only for remediation but in prevention of reading problems. The study by Jansky and de Hirsch (1972) indicated that teachers judged to be competent by their principals had half the number of high risk children who failed to learn to read by the end of grade two as compared with those judged to be incompetent. An earlier study by Chall and Feldmann (1966) found that first grade teachers rated excellent by regular observers made a significant difference for the better in the reading achievement of their first graders.

All of this brings back an earlier time during the 1950s when some classroom teachers were afraid to teach reading to children who had emotional problems. As time passed, it was realized that not only does teaching result in no harm, but when done well it becomes therapeutic (Dahlberg et al., 1952). So, too, for the child with a reading disability, dyslexia, or a learning disability, the research during the past decade suggests that probably the best form of treatment is good teaching by a qualified special remedial teacher and by a good classroom teacher.

Footnotes

1. The use of the term *learning disability* has both positive and negative features. On the positive side is its inclusion of a wider range of school learnings such as math, science, and social studies as subject to specific difficulty in spite of normal intelligence. On the negative side is the global nature of the term. When a teacher or a parent is informed in a report that a child has a learning disability, it is not clear what is meant in terms of the school curriculum. According to referrals to the Learning Disability Clinic of the Children's Hospital Medical Center, Boston, about 75 percent of the children referred to the Clinic have reading disability. That is, most children

referred for an evaluation to this learning disability clinic had a presenting problem of reading difficulty.

2. This is reminiscent of some of the earlier methods' studies whose results were not in line with prevailing philosophical views (see Chall, 1967).

References

Allington, Richard L. "Sticks and Stones . . . but, Will Names Never Hurt Them?" *The Reading Teacher*, 28 (January 1975), 364-369.

Bannatyne, Alex. "Programs, Materials, and Techniques: Review of 'Power in Perception for Young Children,' " *Journal of Learning Disabilities*, 7 (June-July 1974), 333-343.

Bateman, Barbara. "The Role of Individual Diagnosis in Remedial Planning for Reading Disorders," *Reading Forum*, NINDS Monograph 11. Bethesda, Maryland: National Institute of Neurological Diseases and Stroke, 1971.

Bateman, Barbara. "Teaching Reading to Learning Disabled and other Hard-to-Teach Children," in Lauren Resnick and Phyllis Weaver (Eds.), *Theory and Practice in Early Reading*. New Jersey: Laurence Erlbaum, in press.

Black, William F. "The Word Explosion in Learning Disabilities: A Notation of Literature Trends 1962-1972," *Journal of Learning Disabilities*, 7 (May 1974), 323-324.

Black, William F. "Achievement Test Performance of High and Low Perceiving Learning Disabled Children," *Journal of Learning Disabilities*, 7 (March 1974), 178-182.

Burns, Gary W., and Billy L. Watson. "Factor Analysis of the Revised ITPA with Underachieving Children," *Journal of Learning Disabilities*, 6 (June-July 1974), 371-376.

Bryan, Tanis H. "Learning Disabilities: A New Stereotype," *Journal of Learning Disabilities*, 7 (May 1974), 304-309.

Chall, Jeanne S. "Clinical Studies Useful to the Reading Specialist," in Mildred A. Dawson (Ed.), *Combining Research Results and Good Practice*, 1966 Proceedings, Volume 11, Part 2. Newark, Delaware: International Reading Association, 1967, 83-89.

Chall, Jeanne S., and Shirley S. Feldmann. "First Grade Reading: An Analysis of the Interactions of Professed Methods, Teacher Implementation, and Child Background," *The Reading Teacher*, 19 (1966), 569-575.

Chall, Jeanne. *Learning to Read: The Great Debate*. New York: McGraw-Hill, 1967.

Chall, Jeanne S., and Allan F. Mirsky (Eds.). *Education and the Brain*, NSSE Yearbook. Chicago: Chicago University Press, 1978.

de Hirsch, Katrina, and Jeannette Jansky. *Predicting Reading Failure*. New York: Harper and Row, 1966.

Hammill, Donald D., and Stephen C. Larsen. "The Relationship of Selected Auditory Perceptual Skills and Reading Ability," *Journal of Learning Disabilities*, 7 (1974), 429-435.

Hurwitz, Irving, et al. "Neuropsychological Function of Normal Boys, and Boys with Learning Problems," *Perceptual and Motor Skills*, 35 (1972), 387-394.

Jansky, Jeannette, and Katrina de Hirsch. *Preventing Reading Failure*. New York: Harper and Row, 1972.

Masland, Richard L., and Bryant J. Cratty. "The Nature of the Reading Process, the Rationale of Noneducational Remedial Methods," *Reading Forum*, NINDS Monograph 11. Bethesda, Maryland: National Institute of Neurological Diseases and Stroke, 1971.

Niensted, Serena. "Talking with Learning Disability Teachers," *Reading Teacher*, 28 (April 1975), 662-665.

Richardson, Ellis, and Michael Bradley. "ISM: A Teacher-Oriented Method of Reading Instruction for the Child-Oriented Teacher," *Journal of Learning*

Disabilities, 7 (1974), 344-352.

Richardson, Ellis, Bernard G. Winsberg, and Irving Bailer. "Assessment of Two Methods of Teaching and Phonic Skills to Neuropsychiatrically Impaired Children," *Journal of Learning Disabilities*, 6 (1973), 628-635.

Saphier, Jonathan D. "The Relation of Perceptual-Motor Skills to Learning and School Success," *Journal of Learning Disabilities*, 6 (1973), 583-591.

Silberberg, Norman E., Iver A. Iversen, and Jean T. Goins. "Which Remedial Reading Method Works Best?" *Journal of Learning Disabilities*, 6 (1973), 547-556.

The ability to hear the separate sounds in words—or as Walter MacGinitie says, the ability to segment speech into units—is an important component in most reading programs. Unfortunately, there are a number of beginning readers who have difficulty with this task. What can be done to help these students? Without suggesting that a foolproof solution is being offered to overcome these problems, this article does explain why this task is so difficult to master and sets forth a number of solid suggestions which the teacher may use to help these students.

Children's understanding of linguistic units[1]

WALTER H. MACGINITIE
Teachers College
Columbia University

Whatever system of reading instruction is used, children who learn to read eventually learn how to work out for themselves plausible pronunciations of words they have not read before in print. Even methods that do not emphasize decoding strategies at the outset eventually aim to produce readers who can decode in the sense that the term is used in reading instruction. Children who learn to decode must learn, at least implicitly, the properties of a great many linguistic units. Furthermore, they must learn to manipulate these units in various rule-governed ways. They must learn how to recognize, analyze, combine, and recombine English phonemes, syllables, words, phrases, and sentences. On the graphic side, they must learn to work with letters (both capital and small), syllables, words, phrases, and sentences.

"Manipulate these units" refers to such processes as the following: children must learn that the sound represented by the letter *t*, when it occurs at the end of a syllable, is perceived as similar to the sound represented by that letter when it occurs at the beginning of a syllable; they must learn that the same string of printed words has a different intonation contour just before a period and just after a period; they must learn that the root spelled *f-o-r-m* has the same pronunciation in *deforḿ* and *reforḿ* and a related pronunciation in *úniform*, and that each of these words has a semantic relationship with the root. Concerning *spoken* language units, children who are learning to read already have a great capacity, of course, for understanding and manipulation, although they will increasingly encounter written structures that are unlike anything they are likely to use in conversation.

Sources of difficulty in rule learning

One conclusion that has become very clear from accumulated research is that children have a great capacity for learning to manipulate elements of a task according to whatever rules seem to produce sensible outcomes. A companion finding, however, is that it is much harder to learn to manipulate units according to productive rules whenever the task makes other heavy demands. Among the types of task demands that make it harder to learn and apply rules are: 1) heavy memory load, 2) units that are hard to isolate, 3) units that change value depending on some other condition, and 4) the presence of some units that freely change value (French, 1976). That is, although children have the basic ability—one might almost say, the basic urge—to work out simple rules, certain kinds of tasks make it difficult for them to do so. Reading is the sort of task that presents particular difficulties in learning and applying rules, in spite of the child's urge to find rules or reasons. Some of the ways in which learning to read presents these four obstacles to rule learning are described below.

Heavy memory load. Learning rules for decoding English is undoubtedly made much more difficult by the fact that there are 26 letters and more than 40 segmental phonemes rather than 5 or 6 of each. It is hard for children to figure out the appropriate grapheme-phoneme relationship when they need to discriminate between a great many letter forms and remember even the most frequent phonemic value represented by each of these forms. Added to this complexity is the fact that many letter combinations (e.g., *ng, ph, ough*) also represent characteristic phonemic values.

Units that are difficult to isolate. Even syllables are sometimes difficult to isolate. If there were some little mark between printed syllables that corresponded to spoken syllables, children should find decoding easier. Children often must be hampered by confusing instructions from us adults when we confuse our conventions for dividing *printed* words into syllables (as when we hyphenate them at the end of a line) with natural division of words into spoken syllables (Zuck, 1974). Phonemes represent a much more difficult problem, as will be seen later. In any case, it is hard to learn rules involving the recombination of units unless the units are at least implicitly identified.

Units that have values that are conditional upon their context. Children have already learned many conditional values of speech sounds. By "conditional values of speech sounds," is meant such relations as the following: In English the initial /p/ as in *pin* is usually aspirated. You can feel the puff of air that corresponds to this aspiration. When the /p/ follows /s/, as in *spin*, however, it is often not aspirated. You may not feel a puff of air from the /p/ in *spin*. Children do not know these facts in the sense of being able to

tell you about them, but then neither do most adults, and both children and adults have learned what we might call this rule of speech. These conditional relations in *speech* give children little problem in reading.

The only time these conditional pronunciations are likely to cause difficulty for children is when the teacher gives them misinformation based on printed correspondences. For example, there is some dispute as to whether that /p/ in the word *spin* should be a /p/ at all. The problem arises from the fact that English does not maintain a voiced/voiceless contrast of stops following /s/. That is, we do not have a word like *sbin* in English. Some linguists have even suggested that the /p/ following /s/ is more like a voiced phoneme than like a voiceless one, and that perhaps we should consider the labial stop in that position as an entirely separate phoneme. Indeed, young children who are just learning to spell are more likely to use the letter *b*, right after *s*, thus spelling *spin*, *s-b-i-n* (Fink, 1974). Children will have no trouble learning to spell such English words as *spin* or *span* according to our convention for spelling them if they are given time and practice. But too many adults will correct the child who spells *spin* as *s-b-i-n* by saying, "Listen to the way it sounds—*spin*— can't you hear the *p*?" An honest (and precocious) child would say, "No, and an acoustic analysis doesn't disclose it very well either." This is a minor example of an important general principle: We must not assume that words are pronounced the way we adults think they should be, based on the way they are spelled.

Although children will already have learned many conditional relations in speech, they must quickly learn new conditional relations when they begin to learn to read. An obvious example is the well-known problem of learning the sounds that correspond to *c* when *c* is followed by *a*, *o*, or *u* and when *c* is followed by *i* or *e*. These conditional relationships are hard to learn compared to consistent relationships (French, 1976). They are particularly difficult if other demands on memory are simultaneously large, as is typically the case in learning to read.

Variable relationships. The problem of variable relationships is most easily seen, and most familiar to us, in the problems of learning to decode vowels. We are all familiar with the very difficult problem that is presented to children by the fact that the double *o* in *door* is not pronounced like the double *o* in *foot*, that the *u* in *but* is not pronounced as the *u* in *put*, that the *a* in *cat* is not pronounced as the *a* in *father*. In addition to this well-known aspect of the vowel problem, there is the fact that vowels are simply not pronounced and perceived in such neat, clear categories as most consonants are. Vowels are likely to be more variable in both our production of them and in our perception of them (Liberman, 1973).

Analyzing language

In recent years, a great deal has been made of the fact that children in kindergarten or first grade are already quite accomplished in their abilities to manipulate the units of spoken language. Learning to read, however, presents children with a serious *new* task in understanding *spoken* language. This new task is not that of manipulating the units of language, but that of *analyzing* spoken language into units of different types and sizes (Liberman, 1971).

Words as units. As a simple example, suppose children are being taught by a whole-word method. In order to learn to read by this procedure, they must learn that certain configurations of printed letters correspond to particular spoken words. Many investigations have now made quite clear that children's conceptions of what spoken words are—that is, where particular words begin and end in the stream of speech—are rather nebulous (e.g., Holden and MacGinitie, 1972). Children are likely to combine into a single word what we would think of as more than one word. They are apt to regard some syllables or bound morphemes that we consider only part of a word as being whole words in themselves, and they are likely to be inconsistent in specifying word boundaries—at least inconsistent with respect to our printing conventions. For example, in some sentences they will regard *of* as a separate word, and in other sentences they may believe that it is just part of a larger word, such as *ofthe*. This requirement for a *conceptual analysis* of language may be a serious obstacle to learning to read, and we must study and understand this requirement if we are to make basic strides in reading instruction.

Although children's ideas of spoken word boundaries are vague and often do not correspond to what we write as separate words, it does not necessarily follow that children's admittedly vague concepts of words are a fundamental problem in learning to read (Evans, 1975).

Syllables as units. Proceeding to a more molecular level, the problem of analyzing speech into *syllables* does not seem to be a serious block to learning to read. Most six-year old children can learn, without much difficulty, to divide spoken words accurately into syllables (Liberman et al., 1974). Dividing *printed* words into syllables that clearly relate to the separate spoken syllables is not always easy. Words like *already* and *emphasize* can be quite puzzling to children who are trying to break them up into pronounceable syllables, and a complete set of rules for dividing printed words into syllables is confusingly complex (McFeely, 1974). So, finding printed syllables that correspond to spoken syllables *can* sometimes be a real problem. Nevertheless, children can make a lot of progress in the early stages of reading without knowing anything very complex about how to break printed words into syllables. In the

first place, the great majority of words that children read have only one syllable. Fry's Readability Graph (1968) shows that, in second grade materials, words average only 1 and 1/10 to 1 and 1/3 syllables per word. That is, less than 1 word in 3 has more than one syllable. Not until the tenth grade level do typical materials average even 1 and 1/2 syllables per word. In the second place, most of the words that do have more than one syllable have only two syllables, and many of those are easy to divide.

Phonemes as units. Proceeding to a still smaller unit, the segmental phoneme, and to grapheme-phoneme correspondences, we encounter what research now seems to show to be a fundamental problem in analyzing language into units and a basic obstacle in learning to decode. Much of the remainder of this paper will describe some of the evidence that shows why segmenting speech into phonemes is difficult (or, in fact, generally possible only in a conceptual rather than a physical sense). The final section of the paper will discuss some possible implications of that evidence.

Segmenting speech[2]

In order to understand the problem of segmenting speech into phonemes, it will first be necessary to understand the general characteristics of speech as shown on a spectrogram. A speech spectrogram is a visual representation of the component sounds of speech. The first five figures that follow will be used to show the nature of speech spectrograms. These five figures are all from *The Speech Chain* by Denes and Pinson (1963/1973).

MATERIAL REMOVED DUE TO COPYRIGHT RESTRICTIONS

Time in Seconds

Figure 1: Speech spectrogram of a sentence. (From *The Speech Chain* by P. B. Denes and E.N. Pinson. Copyright (c) 1963 by Bell Telephone Laboratories, Inc. Reprinted by permission of Doubleday & Company, Inc.)

Figure 1 shows a spectrogram of a simple sentence, "I can see you." The spectrogram reads from left to right. The dark blotches represent acoustic components of speech sounds. You can see there

are four columns of these dark blotches with relatively clear spaces between them. The four dark columns represent the four words with junctures between them. Thus, time moves from left to right. The vertical dimension shows the sound frequency in Hertz (cycles per second). Dark blotches near the bottom represent low-pitched sounds—sounds of low frequency. Dark blotches near the top represent high-pitched sounds—sounds of high frequency. As you can see, any speech sound is a composite of several sounds in different frequency ranges.

Figure 2 shows some individual words in more detail. There are three separate words: *card, tool,* and *seal.* In the middle of the first word, *card,* you see several dark, horizontal bands. These are called *formants* and are the defining acoustic features of the vowel in the middle of the word *card.* These straight horizontal bands represent the vowel. The consonants before and after the vowel are represented by the upward and downward curves of the bands and by other marks. The /c/ at the beginning of the word is not so dark; it is an unvoiced consonant. The last word, *seal,* begins with a burst of noise representing the hissing of its initial phoneme. This burst of noise covers the entire frequency range of the spectrogram, but is concentrated in the upper portion.

Figure 3 shows two steady-state vowels (two pure vowels), /æ/ (as in *had*) and /a/ (as in *father*). The second one looks roughly similar to the vowel in *card* in Figure 2. It is clear that the differences between the two vowels in Figure 3 are primarily in the differing placements of the formants—the different heights and relationships among the straight, dark, horizontal bands. The relationship among the formants is easiest to see in the first vowel, /æ/ (as in *had*)). If you were to produce the three relatively pure sustained tones at the frequencies shown by the dark bands and play these three tones simultaneously, the combination would sound quite like the vowel /æ/. Actually, the two bottom formants (the first and second) would be sufficient to produce a reasonably recognizable /æ/.

Figure 4 shows two diphthongs—changes from one vowel to another. The first is /ai/ as in *might.* You can hear the vowel change if you say the word *might* slowly, and you can see the corresponding change on the spectrogram. The other diphthong is /au/ (as in *shout*).

Such spectrograms are made by a special recording apparatus that changes sound energy at different frequencies into marks on a moving film at different distances from the bottom of the film. It is possible to reverse this process. It is possible to build a machine that will read spectrograms and play back all the indicated simultaneous sounds as the film passes through the machine. This pattern playback machine produces speech that may sound a bit mechanical but that is readily understood. By changing the spectrographic speech patterns in systematic ways and then running them through a pat-

MATERIAL REMOVED DUE TO COPYRIGHT RESTRICTIONS

Time in Seconds

Figure 2. Speech spectrograms of three individual words. (From *The Speech Chain* by P. B. Denes and E.N. Pinson. Copyright (c) 1963 by Bell Telephone Laboratories, Inc. Reprinted by permission of Doubleday & Company, Inc.)

MATERIAL REMOVED DUE TO COPYRIGHT RESTRICTIONS

Time in Seconds

Figure 3. Speech spectrograms of two steady-state vowels. (From *The Speech Chain* by P.B.Denes and E.N. Pinson. Copyright (c) 1963 by Bell Telephone Laboratories, Inc. Reprinted by permission of Doubleday & Company, Inc.)

MATERIAL REMOVED DUE TO COPYRIGHT RESTRICTIONS

Time in Seconds

Figure 4. Speech spectrograms of two diphthongs. (From *The Speech Chain* by P.B. Denes and E.N. Pinson. Copyright (c) 1963 by Bell Telephone Laboratories, Inc. Reprinted by permission of Doubleday & Company, Inc.)

tern playback machine, it is possible to study the function of various aspects of the acoustic signal. In other words, it is possible to delete one of those dark blotches on the spectrogram, or change its shape, and see what that change does to the perception of what is being said.

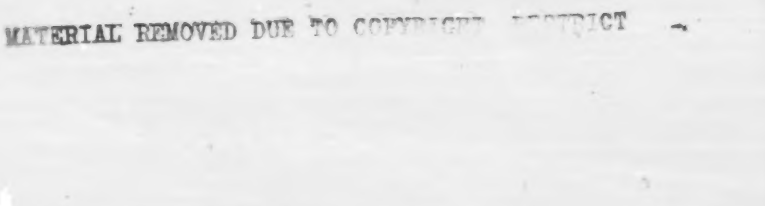


Figure 5. Top: spectrogram of a naturally produced sentence. Bottom: painted pattern that would sound like the same sentence as the sentence at the top when played on a pattern playback machine. (From *The Speech Chain* by P.B. Denes and E.N. Pinson. Copyright (c) 1963 by Bell Telephone Laboratories, Inc. Reprinted by permission of Doubleday & Company, Inc.)

Figure 5 shows, at the top, a spectrogram of a complete sentence and, at the bottom, a simplified reproduction of the same sentence. The sentence, "Many are taught to breathe through the nose," has frequently been used as a pictured example because it is composed of speech sounds that were among the first to be studied and analyzed. Look at the simplified reproduction below the spectrogram. The relatively smooth heavy marks in this simplified version represent the essential parts of the original spectrogram at the top. They were painted on a band of film so they could be run through the pattern playback machine. If the original spectrogram at the top and the simplified, painted version below it were both run through the pattern playback machine, they would sound very much alike. If you examine the original spectrogram and the simplified one, you will be able to detect many correspondences. It is also clear that many features in the original spectrogram have been left out of the simplified one. Those features that remain in the simplified spectrogram are important acoustic features for perception of this sentence.

Now that the reader knows what the essential acoustic features of speech look like on a spectrogram, the problem of segment-

ing speech into phonemes can be considered. Figure 6 shows simplified spectrograms of two open syllables. The figure is from the article by Liberman et al. (1967) that first brought to the widespread attention of behavioral scientists the complex manner in which phonemes are encoded in the stream of speech. If the simplified patterns shown in Figure 6 were run through a pattern playback machine, the listener would hear a voice say /di/ (as in *deed*) and then /du/ (as in *do*). These are much simplified patterns, but they are sufficient for the perception of /di/ and /du/. The thick, straight, horizontal parts of the pattern for /di/ represent the vowel /i/. These horizontal segments are the first and second formants for the vowel /i/. In the pattern for /du/, the thick, straight, horizontal parts represent the first and second formants of the vowel /u/. These horizontal parts of the two syllables, if run through the pattern playback machine, would be heard as the vowels /i/ and /u/.

MATERIAL REMOVED DUE TO COPYRIGHT RESTRICTIONS

Figure 6. Spectrographic patterns sufficient for the synthesis of /d/ before /i/ and /u/. (From "Perception of the Speech Code" by A.M. Liberman, F.S. Cooper, D. Shankweiler, and M. Studdert-Kennedy, *Psychological Review*, 1967, *74*, 431-461. Copyright 1967 by the American Psychological Association. Reprinted by permission.)

The two open syllables in Figure 6 both begin with the stop consonant /d/ (as in *dog*). The question now is, where is the /d/ in each of the two syllables in Figure 6? Clearly, the /d/ must have something to do with those curved portions at the beginnings of the pairs of horizontal bars. These curved portions are called transitions; the lower one in each case is called the first formant transition, and the upper one is called the second formant transition. Since both syllables begin with /d/, we naturally look for the /d/ in the part that is the same in the two spectrograms. The first formant transition, the curved portion leading into the lower bar, looks the same in both spectrograms. In fact, it is the only feature at the beginning of the two syllables that does look the same. But it is not the /d/. If the first formant transition were run through the pattern playback machine, it would sound like a low, rising whistle. In fact, Liberman et al. (1967) point out that this particular first formant transition is the same for all the voiced stops. It would also look like this if the paired syllables were /bi/ and /bu/ or /gi/ and /gu/.

MATERIAL REMOVED DUE TO COPYRIGHT RESTRICTIONS

Figure 7. Spectrographic patterns sufficient for the synthesis of /d/ before various vowels. (From "Perception of the Speech Code" by A.M. Liberman, F.S. Cooper, D. Shankweiler, and M. Studdert-Kennedy, *Psychological Review*, 1967, *74*, 431-461. Copyright 1967 by the American Psychological Association. Reprinted by permission.)

The portion of these spectrograms that *is* distinctive for /d/ is, in fact, the *second* formant transition. It does vary, however, depending on the following vowel. Figure 7 shows five open syllables that all begin with /d/. The second formant transition changes systematically, and none of these second formant transitions, heard in isolation, sounds anything like /d/. They sound like rising or falling whistles and not like speech at all. Thus, there is no specific feature or even any stable combination of features that we can say represents /d/. This is the first important principle: the acoustic features that are recognized as a particular stop consonant phoneme change, depending on the vowel that follows. This generalization holds also for most other consonants. Vowels that precede consonants also influence the acoustic form of the consonants. To quote Liberman et al. (1967), "It is the usual case that the acoustic cues for a consonant are different when the consonant is paired with different vowels . . . [and] when it is in different positions with respect to the same vowels . . . " (p. 440). Thus, although we perceive the /d/ in the two syllables shown in Figure 6 as the same phoneme, the acoustic signal that produces that perception is very different. Where is /d/? Nowhere in particular.

There is another sense in which /d/ is nowhere in particular. It is not possible to cut off the front of either of these spectrograms at some point and obtain a signal that would be perceived as an isolated /d/. If just the left hand tip were cut off and run through a pattern playback machine, it would not sound much like speech at all. If more and more of the lefthand part of one of these spectrograms were included, it would continue sounding like a nonspeech sound, till suddenly it would sound like /d/ *followed by a vowel.* There is no portion of the spectrogram that represents /d/ alone. The basis for this second principle is that the speech sounds, as portrayed in the spectrograms, provide information about two phonemes at once—the consonant and the following vowel.

It is *generally* the case that successive phonemes are merged together like this in the sound stream. The sound signal thus cannot usually be divided on the time axis into phonemic segments. At any one instant in time, the sound signal is usually giving information

about two or three phonemes, not just one (Liberman, 1971). This evidence does not mean that phonemes are a useless abstraction. It is still appropriate to conceive of speech partly in terms of a sequence of segmental phonemes. Nevertheless, it is important to know that the acoustic signals for these phonemes usually overlap, so that speech itself generally cannot be cut into separate phonemic segments.

This overlapping of phoneme cues relates to the fact that the human speech apparatus could not produce separate and distinct oral gestures at the rate required for normal rapid speech. Speaking at a rate of 200 words per minute is a rate of about 15 phonemes per second. Neither could the ear identify totally separate sound patterns at that rate (Liberman et al., 1967). Thus, phonemes are produced and processed in parallel. For example, in saying the syllable /di/, the oral cavity for producing the vowel /i/ is already being formed while the /d/ is being produced. In general, the sound that we produce when we utter a given phoneme is influenced by the phonemes that are to follow, for our speaking apparatus is already being rearranged for uttering those following phonemes.

Possible relationships to reading instruction

The preceding section introduced two fundamental principles of speech perception: 1) the acoustic signal that represents a particular phoneme usually depends on the other phonemes that surround it, and 2) many phonemes in the stream of speech cannot be isolated as particular, identifiable parts of spoken syllables or words. What do these two principles have to say for reading instruction?

An earlier section described several conditions that usually make it hard for children to learn and to apply rules. One of those conditions occurs when the units to be manipulated or ordered by the rules are hard to isolate. We certainly encounter that condition when we try to teach phonics rules to children. Many teacher's manuals for beginning reading have the teacher ask children to analyze a word into its component phonemes and then recombine them or substitute for one of them. It is easy enough for literate adults to say that *cat* is composed of /c/, /æ/, and /t/. Literate adults have already developed the conceptualizations that allow them to represent complex intertwined sound signals as a simple series of abstract elements. But it should now be clear how very abstract those phonemic elements, and indeed the whole process, must be for young children.

It is not any easier if the teacher's manual comes at it the other way and tells the teacher to ask the children to synthesize rather than analyze words. That is, it is also a difficult and highly abstract task to build up words from the sounds that supposedly correspond to letters. The sound that corresponds to a letter varies,

depending on the other sounds around it, even though it is *perceived* as being fairly constant. The children will need to take the following vowel into account if they are to pronounce the initial consonant as part of a syllable that includes that following vowel. *Cat*, if pronounced /kə/ + /æ/ + /tə/, is three syllables. How are the children to pronounce three syllables as one? It would take a set of abstract deletion rules to do it. To ask children to make *cat* out of /kə/+ /æ/ + /tə/ is like asking them to solve a jigsaw puzzle in which the pieces do not fit together. Some of the pieces must be slid partly underneath other pieces in order for the final picture to appear. In addition each piece of the puzzle already has a separate and complete picture on it, so it is hard for the children to know which parts of which pictures must be obscured in order for a new, different, larger picture to emerge.

Some procedures for teaching decoding that are already widely used reduce, somewhat, the problems of blending that these facts about the structure of speech illustrate. When words are taught as "families" (*bat, cat, fat, hat, mat*) the children learn the final part of the syllable (*at*) as a unit and must merge only two syllables (e.g., /kə/+ /æt/) to produce a CVC word. In some other approaches to decoding, the first part of a CVC word is learned as an open syllable unit (e.g., /kæ/) and it is different *final* consonants that must be added. There are still two syllables to merge, but probably the more important merger has already occurred. One possible disadvantage of this open-syllable procedure is that, later on, as two-syllable words are encountered, the vowel in the CV syllable will characteristically correspond to a different sound from the value the children have previously learned. Thus, *c-a* may not be /kæ/, but /ke/ as in *caper*.

One other procedure for overcoming the "blending" problems that are presented by the structure of speech involves combining the two partial syllabaries of the "word families" and "open syllables" approaches. That is, the child is taught to recognize and pronounce syllables like *c-a*, /kæ/, and *a-t* /æt/, and then taught to combine them: /kæ/ + /æt/, /kæt/. The possible advantages of this alternative are that a much smaller number of syllables would need to be memorized than for a full syllabary, and combining or blending should be easier. We need to study empirically just how much easier, if any, blending is with this procedure.

A more radical suggestion stemming from these facts of speech structure is to teach written English as a syllabary. Thus, the three letters *b-a-t* would be learned as a syllable pronounced /bæt; *c-a-t* would be learned as a different syllable pronounced /kæt/; *c-u-t* would be leaned as a still different syllable pronounced /kʌt/; and so on. Some preliminary investigations have indicated that such correspondences between spoken and written syllables may be relatively easy to learn compared to the process of learning letter-

phoneme correspondences and then blending them (Gleitman & Rozin, 1973). The principal disadvantage of teaching a syllabary appears to be that much of the economy of an alphabetic writing system is wasted. There are approximately 9000 different one-syllable words in English (Moser, Dreher, and Oyer, 1957). That is a lot of words to memorize if the child is asked to learn each one of them separately without taking advantage of the alphabetic principle, and indeed no one has suggested such wholesale memorization of syllables. Gleitman and Rozin (1973) propose teaching only 60 syllables, chosen to be as productive of other words when combined and as useful in writing as possible.

Is is likely, of course, that once children have learned to name a fair number of syllables by sight, they may then learn grapheme-phoneme correspondences and blending more readily. The overall economy in learning that might be achieved by such a combination of procedures has not really been studied. We should notice, however, that what we might think of as a highly experimental program involving learning a stock of syllables and then turning to the development of more detailed decoding rules is quite similar, in general outline, to the once common practice of first learning a stock of sight words and then turning to decoding rules. The difference, of course, is in the selection of material to be learned by sight: syllables (including one-syllable words) that can be combined to form many other useful words vs words that are generally useful in communication. It is not yet clear how important a difference this is.

Even before much of the needed research on applications is done, studies of phoneme perception seem to suggest two helpful general strategies for reading instruction. These strategies are based on the fact that the natural perceptual units of speech are typically larger than single phonemes. First, when any child in a phonics program appears to be having consistent trouble learning to blend, a logical thing to try would be to help him master some useful syllables.

The second strategy is more a frame of mind. If we really understand the task that children are faced with, we should be better at helping them with it. Lesson plans in current teacher's manuals frequently ask children to analyze a unitary perception into abstract conceptual elements. Or they ask children to forget all that they know about speech and to believe that /kə/ + /æ/ + /tə/, said fast, is *cat*. We do not yet know the best ways around these problems. Research on speech perception has at least told us that these problems exist and has clarified their nature. Further research on beginning reading should evaluate various possible solutions. The best solutions may prove to be different for different children. In the meantime, understanding exactly what children are up against may make us more understanding of their difficulties, more impressed by their progress, and more intelligent in giving help.

Footnotes

1. The author wishes to acknowledge that many of the points in this paper were developed in discussions with Joyce N. French. In general, the points in this paper are not original with the author, but represent a review of research and related interpretations organized with the aim of helping teachers and reading specialists learn about developments that should clarify our understanding of what learning to read involves. The author is grateful to Lila Gleitman and Isabelle Liberman for reading, and commenting on, a draft of this article. Remaining errors of fact or of interpretation are of course the responsibility of the author.

2. The author wishes to acknowledge the major sources from which he has taken the information on phoneme perception. The primary sources are reports of work done at the Haskins Laboratories and the Bell Telephone Laboratories, the most important being the landmark article, "Perception of the Speech Code" (Liberman et al., 1967). Other major sources are articles by Isabelle Liberman (1973) and by Gleitman and Rozin (1973) and the book, *The Speech Chain*, by Denes and Pinson (1963/1973).

References

Denes, P.B., and E. N. Pinson. *The Speech Chain*. Garden City, New York: Anchor/Doubleday, 1973. (Originally published, 1963.)

Evans, M.C. "Children's Ability to Segment Sentences into Individual Words," in G.H. McNinch and W.D. Miller (Eds.), *Reading: Convention and Inquiry*, Twenty-Fourth Yearbook of the National Reading Conference. Clemson, South Carolina: National Reading Conference, 1975.

Fink, R. "Orthography and the Perception of Stops After S," *Language and Speech*, 17 (1974), 152-159.

French, J.N. "A Study of Cue Salience and Rule Identification in Three Beginning Reading Tasks," unpublished doctoral dissertation, Columbia University, 1976.

Fry, E. "A Readability Formula that Saves Time," *Journal of Reading*, 11 (1968), 513-516.

Gleitman, L.R., and P. Rozin. "Teaching Reading by Use of a Syllabary," *Reading Research Quarterly*, 8 (1973), 447-484.

Holden, M.H., and W.H. MacGinitie. "Children's Conceptions of Word Boundaries in Speech and Print," *Journal of Educational Psychology*, 63 (1972), 551-557.

Liberman, A.M., et al. "Perception of the Speech Code," *Psychological Review*, 74 (1967), 431-461.

Liberman, I.Y. "Speech and Lateralization of Language," *Bulletin of the Orton Society*, 21 (1971), 71-87.

Liberman, I.Y. "Segmentation of the Spoken Word and Reading Acquisition," *Bulletin of the Orton Society*, 23 (1973), 65-77.

Liberman, I.Y., et al. "Explicit Syllable and Phoneme Segmentation in the Young Child," *Journal of Experimental Child Psychology*, 18 (1974), 201-212.

McFeely, D.C. "Syllabication Usefulness in a Basal and Social Studies Vocabulary," *Reading Teacher*, 27 (1974), 809-814.

Moser, H., J.J. Dreher, and H.J. Oyer. *One-syllable Words*. Air Force Cambridge Research Center Operational Applications Laboratory Technical Report No. 41. Columbus, Ohio: The Ohio State University Research Foundation, 1957.

Zuck, L.V. "Some Questions about the Teaching of Syllabication Rules," *Reading Teacher*, 27 (1974), 583-588.

Some people believe that research has had little impact on educational practice. Harry Singer's article indicates there are a number of studies which have exerted an impact on teaching. The focus of his article, however, is on recent research which he believes should be taken seriously and which should make a difference in classroom instruction. Do you agree with Dr. Singer's conclusions based upon the studies he has reported here? How does one determine which research studies are significant?

Research in reading that should make a difference in classroom instruction

HARRY SINGER
University of California at Riverside

Some research in reading has made a difference in reading instruction (Russell, 1961). Other research that should have made a difference, hasn't; while some research that did make a difference, shouldn't have (Singer, 1970).

Conditions determining whether studies will make a difference

Russell believed that studies he cited made a difference because they focused upon significant issues in reading and were based upon simple designs and easily understood statistical techniques. However, the investigations that should have made a difference also fit these criteria. Consequently, additional explanations have to be adduced to understand why these studies did not make a difference. On the list of possible explanations are ideological resistance to research results (Moynihan, 1968), findings contrary to "conventional wisdom" (Chall, 1967), acceptability of only those research findings that are in accord with prevailing maturational-environmental bias (Durkin, 1968), variation in adequacy of dissemination of findings (Chall, 1967), susceptibilities of educational decision-makers to commercial propaganda, and maintenance of vested interest. Although I am aware of the pitfalls between research results and changes in practice based upon them, I am nevertheless hopeful that dissemination of research results with implications clearly drawn for practice will have an impact upon reading instruction.

Reading Research that Makes a Difference

Russell's ten "best" or most widely influential studies had the following consequences.

1. **Thorndike, 1917** A change in the definition and subsequent teaching of reading from a process of just sounding out words to getting meaning from the printed page through reasoning or problem solving processes.
2. **Buswell, 1920; Judd and Buswell, 1922** A shift in emphasis from oral to silent reading and use of different kinds of materials to teach children to read for a variety of purposes.
3. **Gates, 1928** Development and use of concrete materials and techniques for teaching various ways of recognizing words in silent reading.
4. **Gates, 1927** Construction and value of tests for diagnosing reading, providing the basis for the eventual establishment of reading clinics.
5. **Terman and Lima, 1929** Selection of children's literature according to developmental changes in children's interests.
6. **Smith, 1934** Appreciation of the evolution of current reading methods and materials.
7. **Gray and Leary, 1935** Objective measurement of levels of difficulty of reading materials.[1]
8. **Agnew, 1939** Insight into some advantages and disadvantages of phonics instruction.
9. **Strang, 1938; 1942** The foundation of and concern for postelementary instruction in reading.
10. **Gray, 1956** Establishment of the need for adapting reading instruction to the language of a country and appraisal of worldwide problems in literacy.

Research that Should Have Made a Difference

Singer's list of significant research studies in reading and their implications for instruction.

1. **1921, Gates** There is a "useful distinction between 'ability to comprehend' and 'rate of reading.' " Hence, schools should test for speed and for comprehension of reading to diagnose and evaluate student progress.
2. **1957, McCullough** Correlations of comprehension scores for tests of main ideas, details, sequence, and creative reading at the second grade level range from .26 to .50 with a median of .45. Comprehension in primary grades is a function of the type of test administered and comprehension tends to be specific. Therefore, use a comprehension test that has curricular validity.
3. **1939, Gates, Bond, and Russell** Whether or not a variable is predictive of reading achievement, especially at the end of the first year of instruction, depends to some degree upon the common elements among the predictor variable, the evaluation instruments, *and* the method of instruction. Hence, determine whether readiness test fits the program of instruction.
4. **1926, Gates** Perception is not a unitary function. Correlations among perception of numbers, geometric symbols and printed words were quite low, averaging only .35. Hence, training in laterality, tracing lines, matching block designs, or motor activities is not likely to transfer to perception of printed words. What is most effective is training on discrimination and perception of printed words.

5. **1940, Gilbert** Eye-movement behavior is disrupted when good reader tries to follow poor reader, and vice-versa. Hence, reading in a circle, using the lock-eyed method of instruction is inappropriate. Have children *listen actively* while one student is reading aloud to audience of other children and teacher is diagnosing the oral reader.

Research that Made a Difference, but Shouldn't Have

6. **1931, Morphett and Washburne** This study had a significant and long-lasting effect, but shouldn't have because reading readiness does not depend solely on a mental age of 6 years, 6 months.
7. **1886, Cattell** This study also had an unjustifiably widespread impact: configuration, based on research with adults, is not a clue for word identification in beginning readers, especially when words taught in primary grades are short and similar in appearance.

In order for the studies I review to make a difference, I know that teachers or educational decision-makers will first have to become aware of the research and decide to act on the results; then they will have to transform the implications into teaching procedures and plans and, in some instances, persuade other faculty members, supervisors or specialists, and administrators to let them put their research-based instructional changes into practice. If teachers undertake this developmental process, then the research reviewed below will make a difference in reading instruction.

The studies which I believe should make a difference, supplement the research listed previously. I have organized them into a developmental sequence. While I could have included other studies—undoubtedly other researchers could construct their own list of studies that should make a difference—the studies included here concentrate on significant aspects of reading instruction: 1) research on the relation between IQ and reading, 2) pacing instruction, 3) teaching word recognition, 4) type and timing of supplementary instruction, 5) impact of a comprehensive instructional program on reading achievement, and 6) an hypothesis on a schoolwide program for the improvement of reading.

Relationship between IQ and reading

IQ tests have been much maligned, partly because they have not been used and interpreted properly. In the field of reading, researchers have usually only reported that IQ has an increasing relationship with reading as students progress through the grades, for example, as Durkin (1966) has done. She reported that the correlation coefficient between IQ and reading increased steadily from .40 at grade 1 to .79 at grade 5. If these results alone are

considered, they might be misinterpreted. The implication that might be incorrectly drawn from them is that a child has to be bright to learn to read. Indeed, Morphett and Washburne (1931) are frequently quoted, even today, as having proven that a mental age of six years and six months is necessary for beginning reading instruction. Their criterion (based upon research that shouldn't have made a difference) is still applied on a widespread basis, despite demonstrations that a "talking typewriter" can be used to teach three-year-olds to begin learning to read successfully (O.K. Moore in H. Rowan, 1961) or that four-year-olds through the normal range of mental ability from IQ about 70, on up, can learn some sight words at age four (Davidson, 1931). Indeed, the age when instruction can be initiated is a function of a number of variables, including a criterion of what constitutes successful instruction over a given period of time and a student-teacher ratio that is correlated with age of the students (Holmes, 1962). Indeed, Gates and Bond (1936) judged that the age for initiating reading instruction was not a function of a biological (maturation) nor a psychological (readiness) criterion but, instead, was determined by a social criterion. Thus, if society considered that it is socially useful to begin teaching children under age five to learn to read, we would do so. Moreover, research has indicated we could do so, provided, of course, the criterion for success is changed and a lower teacher-student ratio is funded.

In fact, IQ for a group of children in the normal range of mental ability (IQ about 70 and above) has a *decreasing* relationship with learning to read, while IQ for the same group of children has an *increasing* relationship to reading achievement or ability to learn from text as the group progresses through school (Singer, 1977a). The reason for this seemingly paradoxical relationship is that reading has two overlapping phases of development which we shall refer to as 1) learning to read or reading acquisition behavior and 2) learning from text or gaining information from text (Singer et al., 1973-1974; Singer and Rhodes, 1976; Singer and Donlan, 1978).

Learning to read consists of acquiring responses to printed words and learning to integrate these responses with previously developed language processes, particularly semantic and syntactic processes. This phase of reading, consisting of close-ended objectives that require convergent thinking, can be mastered by 95 percent of the population (Bloom, 1971), probably by a large percent of the student population prior to grades six to eight. As a group of students begins to master reading acquisition behavior (or, indeed, any phase of instruction), the group's range of individual differences in the particular phase of instruction decreases. Hence, the correlation between achievement in the particular phase of instruction and IQ scores decreases.

The second phase of reading goes beyond the first phase. It also includes use of "materials and processes of thinking" (Russell,

1956) to gain information from text. The materials of thinking consist of percepts, images, memories, associations, and concepts, while the processes of thinking involve discrimination, abstraction, generalization, concept formation, inference, interpretation, problem-solving, and critical and creative thinking. These materials and processes of thinking are open-ended and are never mastered. Consequently, the range of individual differences on these cognitive abilities and processes increases as a group of students goes through school because individuals differ in their rate of learning and level of cognitive capability. Moreover, reading achievement tests are made more difficult from grade to grade, mostly by using more complex syntax and more abstract vocabulary. IQ tests also become weighted with more abstract vocabulary at successive age levels. Consequently, these two tests become more alike. As they do so, it is not surprising to find their correlations increasing at successive age levels.

However, we must try to measure these two aspects of reading separately and keep their implications independent in our thinking and educational planning, even though we cannot completely separate the two components. Nevertheless, we can better approximate their separate assessment by repeatedly using the achievement test administered at the end of each grade, starting with grade one, until students have demonstrated mastery of the test at that grade level. We can also continue to administer reading achievement tests standardized for each grade. The repeatedly administered grade level test can then be called a test of learning to read. The results of repeated testing will demonstrate that students are learning to read and that they are progressing towards mastery of this phase of reading. Typical year-end reading achievement tests can be referred to as tests of ability to read and learn from text or printed materials. Scores on this type of test will reveal how well students, in comparison with their peer group, are able to gain information from text. The results will also show that the range of individual differences in reading achievement or ability to read and learn from text increases through the grades (Singer, 1977b).

Figure 1 depicts a plan for measuring changes in individual differences that occur in both phases of reading development as students progress through the grades. Note that as students master reading acquisition behavior at one grade level, they can go on to repeated testing at another grade level until they demonstrate mastery of reading acquisition behavior at that grade level, too. The implications of these two types of measurement are:

1. All students in the normal range of intelligence can learn to read, given adequate instruction and time for learning. Therefore, schools should allow for varying rates in mastery of reading acquisition behavior through repeated testing and continuity of instruction over the grades.

2. Students differ increasingly in their ability to gain information from text as they progress through the grades. Furthermore, this increase in variability is normal. Consequently, teachers should adapt instruction to this increasing range of individual differences.

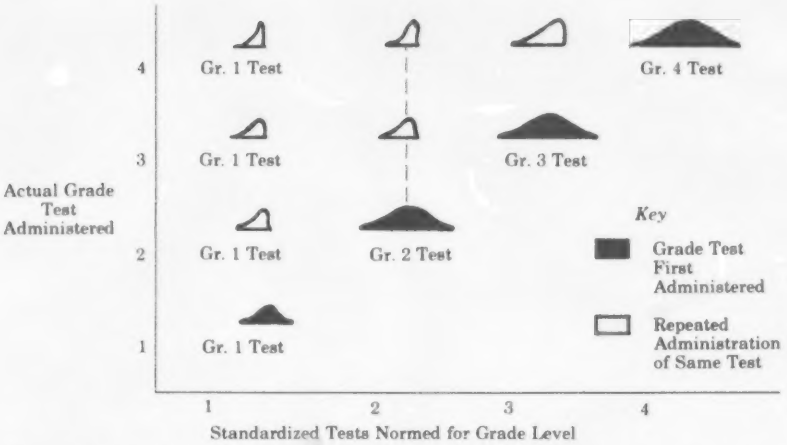


Figure 1. Plan for Assessing Development of Reading Acquisition Behavior and Ability to Gain Information from Test. The figure shows that the standardized test for grade 1, which is repeatedly administered in grades 2, 3, 4, and 5, decreases in variability as increasing numbers of students master it,while grade level standardized tests administered at the end of each grade level (diagonal curve) increase in variability across the grades as students differ more and more in progressive grade levels on their ability to learn or gain information from printed materials.

If schools do initiate repeated testing, we would find that teachers, students, parents, administrators, and the community in general would realize that students are mastering reading acquisition behavior while also increasing in variability of ability to gain information over the grades. Therefore, these two measures could serve as suitable indices for accountability of schools and, perhaps, resolve a perennial conflict between what the school can accomplish and what the community expects and even demands that schools achieve. Schools could reasonably be held accountable for *decreasing* variability in reading acquisition behavior and *increasing* variability of students to gain information from text.

Pacing reading instruction

Since students vary in rate of acquisition in both phases of reading development, teachers must learn to pace students appropriately. Barr (1972) reported data, however, which indicated that the two schools in her study varied in the rate at which students were paced in reading instruction, even though the students appeared to be comparable in socioeconomic level and, presumably, in ability.

An implication from Barr's study is that rate of learning to read is partially a function of how teachers pace their students. Consequently, expectancy criteria, to determine appropriate pacing rates, should be devised. Although such expectancy criteria have been developed and are in use, they are not infallible. California, for example, employs a complex set of criteria (Law, 1975). Moreover, they do not indicate the range of individual difference in expectancy levels within a classroom. Consequently, teachers must also depend upon their own observations, including use of criterion-referenced tests and their own judgment to determine adequacy of instructional pacing of students. The criterion should be the maximum rate at which students can learn. Through experimentation such as observing the rate at which individual students master lessons (for example, recognition of sight words or acquisition of grapheme-phoneme correspondences), teachers can arrive at their students' most effective learning rates. An example of a more formal type of learning rate test consists of teaching and measuring over several trials a student's acquisition of a group of sight words, grapheme-phoneme relationships, or some other aspec of word recognition (Singer and Beasley, 1970). By applying their knowledge of students' learning rates to pacing students properly, teachers can not only stimulate maximum cumulative reading development but also prevent cumulative reading failure.

Helpful for determining students' rate of learning and also for establishing student motivation for learning is a procedure for putting rate of learning under the control of the student. This procedure, which can be called a *motivational graph method*, consists of having pupils participate in setting their own achievement goals, giving students adequate time to progress towards their goals, and plotting the results on a cumulative graph. This procedure proved to be very effective in motivating a seriously disabled reader and may be even more effective in developmental instruction (Singer and Beasley, 1970).

Word recognition strategies

Important in pacing and motivating pupils is not only rate at which students are taught and amount of instructional time, but also how pupils are taught to recognize printed words. While various

cues can be used to help pupils identify words, including context and picture clues, pupils must eventually acquire responses to the printed words themselves. Singer, Samuels, and Spiroff (1973-1974) have demonstrated that elimination of external cues, such as context and/or pictures, and focusing attention on the printed words themselves is a more efficient procedure for learning sight word recognition. Although students tend to make more rapid initial progress through the use of external cues, the cues are no longer sufficient when similar words appear or when synonyms occur or when students are asked to transfer their word recognition ability to novel words (Samuels, 1976).

In general, what appears to facilitate and affect word recognition ability is training in perception of printed letters, phoneme-grapheme correspondences, functional units such as blends and digraphs, and ability to blend these constituents of a word together. Students must also learn to use their language abilities, including syntactic and semantic processes, for formulating expectancies on word identification in the process of reading (Samuels, 1973). Furthermore, they must also develop automaticity of word recognition, particularly for high frequency words, so that they can concentrate their attention on comprehending or gaining information from text (LaBerge and Samuels, 1976). However, some students who do not make expected progress in classroom instruction alone can frequently benefit from supplemental instruction.

Supplemental reading instruction

As more aides enter the classroom and as more specialists are trained to assist the classroom teacher, the questions asked are: "When should supplemental instruction be given?" and "What content should be used in this supplemental instruction?" The answers can be gleaned from two research studies. Gates and Bond (1936) found that those pupils who were making least progress in reading made "marked" improvement after three weeks of special instruction while still in first grade. They also pointed out that in this supplemental instruction the "techniques and materials of reading can be adjusted to teach children successfully . . ." (p. 685). More recently, Ellson et al. (1968) confirmed Gates and Bond's earlier results that supplemental instruction in the primary grades using the same content taught in the classroom enhances reading achievement. Both of these studies support a theory of school learning which emphasizes, in part, that aptitude is inversely related to time to learn and that quality of instruction is a function of the degree to which tasks have been properly sequenced, presented, and adapted to the learner (Carroll, 1963).

Hence, supplemental instruction that provides students who have lower aptitude with additional time to learn, and paces students properly through the sequence of tasks, will tend to be beneficial. Furthermore, supplemental instruction on materials used in the class can significantly improve the reading achievement of first graders. Since Ellson et al. had used programed tutoring, it appears that the instructional strategies of programed instruction with its careful directions, systematic sequencing, step-by-step feedback, individually determined rate and progress, and variable time for learning, which are consistent with a theory of school learning, are effective supplemental procedures for enhancing reading instruction.

Intervention programs based on a *change* in curriculum do not satisfy the aptitude and the quality of instruction criteria. Instead, they *add* to the learning burden. In short, the best intervention for low achieving students should come during the first grade and should supplement classroom instruction. But in some schools supplemental instruction alone is not enough, particularly where there is significant disparity between average rate of reading development for boys and girls.

Differences between boys and girls in rate of reading development

Boys are usually slower than girls in rate of reading development (Balow, 1963; Singer, 1964), but under certain cultural conditions boys can be superior. For example, in Germany, boys had superior reading achievement to girls (Preston, 1962). However, even in the United States where cultural conditions are presumably favorable to reading achievement for girls, instructional conditions can be devised which eliminate sex differences in rate of reading acquisition. These conditions, constructed by McNeil and Keislar (1964), not only provide strong evidence for a cultural explanation of sex differences in reading achievement but also provide insight into instructional dynamics that lead to sex differences in achievement in American elementary schools which, unlike German schools, are dominated by female teachers.

In their experiment, McNeil and Keislar, using a programed autoinstructional device to teach kindergarteners, found that boys and girls acquired sight word recognition at an equal rate, even though the girls had higher reading readiness scores. But, after instruction in a typical first grade classroom, the girls began to achieve at a faster rate of acquisition. An interview with the boys and girls provided insight into the acquisition gap. To the question, "Who in your group does the teacher most often call upon to read?"

the answer usually was "a girl" but when asked, "To whom does the teacher say 'Sit up!' 'Turn around!' 'Pay attention!'" the answer was most often "boys." In general, girls tended to receive *reading instruction* while boys got *behavioral modification.*

Similar differences in behavior have been reported for high and low reading achievement groups: the high group cooperated with their teacher to maximize reading instruction while the low achievers and their teacher used a variety of "misbehavior" to minimize reading instruction (McDermott, 1976). A possible interpretation is that the students and their teacher were consciously or unconsciously avoiding a difficult and embarrassing situation of learning to read.

The implications from these two studies are clear: 1) create an environment more appropriate to boys, perhaps through use of more male teachers in primary grades; 2) use instructional aides to help manage the class while reading instruction is given; 3) attempt to prevent embarrassment in oral reading, perhaps through emphasis on silent reading before oral reading or through group oral reading; 4) intervene as early as possible, for example, in follow-up lessons for any students not making expected progress, which Gates and Bond (1936) and more recently Ellson et al. (1968) found were salutary provided the follow-up instruction, given by aides or by specialists, supported and reinforced ongoing instruction; and 5) adapt the pace of instruction to stimulate maximum positive and cumulative reading development while minimizing a negative accumulation of errors (Barr, 1972).

In some schools, however, it is not enough to eliminate sex differences through a change in the culture and improve the class's reading development through supplemental instruction. In these schools, the total quality of instruction may have to be improved.

Comprehensive instructional program

Over the short run, a narrowly based instructional program can be successful because the load on learning is reduced and the criterion for achievement is narrower. Consequently, the time to learn can be shorter. Over the long haul, however, the more comprehensive program (which, of course, requires more time to learn) can be more successful because it contains the necessary components for a broader criterion and more mature level of reading achievement. The evidence for this conclusion comes from a two-year, longitudinal study conducted by Ruddell (1968). He compared four programs: programed instruction emphasizing phoneme-grapheme correspondence was compared with a basal reader program stressing, initially, whole word recognition; two other programs consisted of supplementing the programed instruction and basal reader programs with instruction that taught students meaning of printed morphemes and use of syntax in reading. At the

end of the first year, the narrower-based, supplemented programed instruction resulted in superior achievement, but at the end of the second year the supplemented basal reader instruction yielded significantly higher reading achievement.

The supplemented basal reader not only had all the necessary subskills components in its sequence of instruction but also was devised to maximize, in the process of reading, the use of all the subskills taught in the program. Developmental instruction in reading skills is not complete unless students are taught to integrate the subskills into the process of reading. Like the swimming instructor, the reading teacher should analogously direct students: "Now that you have learned this skill, jump into your reading material and use it." More specifically, after first graders have learned a grapheme-phoneme correspondence for an initial consonant, the teacher can direct them to read a story and each time they come across the initial consonant and respond with the corresponding phoneme they should signal silently that they had done so by crooking their finger or some other appropriate silent motion. Thus, students will be intentionally mobilizing their new skill in the process of reading and integrating it into their repertoire of skills for attaining reading achievement.

Although a classroom teacher can make a difference in students' reading achievement in the classroom during the year, what makes a cumulative difference in a school is not the classroom teacher alone but a schoolwide approach to improving reading development. Essentially, we have to modify the hypothesis that the teacher alone makes the difference in reading instruction and substitute a broader hypothesis for improving a school's reading achievement.

Hypothesis for improving reading achievement

Although Ruddell found that a comprehensive program, consistently applied, made a significant difference in reading achievement, the conclusion drawn from the results of all the First Grade Studies is that the variation in reading achievement within a method was greater than the variation in achievement among the methods (Bond and Dykstra, 1967). Since classes taught by some teachers attained high reading achievement while classes with comparable students taught by other teachers had relatively low achievement with the same method in the First Grade Studies—which attempted to compare all methods (e.g., basal vs. programed instruction vs. initial teaching alphabet)—an hypothesis was formulated to explain the results. The hypothesis, which is still untested, is that it is not the method, nor the program, but the teacher who makes the difference in reading achievement. An implication would be that each teacher in a school could adopt any method or program of instruction and yet attain high achievement.

While particular teachers might be successful, if each teacher at successive grade levels in a school used a different program, the overall achievement of the school would not likely be high. The reason for this apparent paradoxical conclusion is that students in the school would not have a cumulative program that would pace them appropriately and provide for their individual differences in aptitude and in time needed for learning. Indeed, in case studies of three schools which had different programs but were alike in faculties trained and committed to the *same* instructional program, reading achievement in all three schools was significantly above average (Singer, 1977c). Hence, the modified hypothesis for schoolwide achievement is that while a teacher can make a difference in reading achievement in a given year, what makes for a cumulative difference in a school's reading achievement is a coherent and systematic program throughout the grades to which the faculty is committed.

Conclusion

In a coherent and systematic program a school can employ repeated testing that is curricularly valid and demonstrate that students are progressing in both phases of reading development. Teachers can then be trained to use techniques that will promote necessary skills required in the program. Those students whose aptitudes require supplemental instruction can receive it from aides or specialists familiar with the school's program. But contrary to conventional wisdom, a comprehensive program over the long haul yields more superior results than does a narrower program. If the hypothesis concerning a coherent and systematic application of a schoolwide reading program is tested, together with the implications drawn from research reviewed above, then we are likely to find that research in reading that should make a difference, will.

Footnotes

1. For a current review of computational readability formulae, see Klare (1974-1975); for a noncomputational technique, see Singer (1975).

2. A more recent study that had a wide impact contained the hypothesis that dialect interfered with comprehension (Goodman, 1965); the hypothesis was subsequently found to be false (Goodman and Buck, 1973). However, a more recent investigation indicated that dialect, only in its syntactic component, may have merely a transitory effect upon oral reading achievement (Lucas and Singer, 1975).

References

Agnew, Donald C. *The Effect of Varied Amounts of Phonetic Drill on Primary Reading*. Duke University Press, 1939.

Barr, Rebecca C. "The Influence of Instructional Conditions on Word Recognition Errors," *Reading Research Quarterly*, 7 (1972), 509-529.

Balow, Irving H. "Sex Differences in First Grade Reading," *Elementary English*, 40 (1963), 303-306, 320.

Bloom, Benjamin S. "Mastery Learning and Its Implications for Curriculum Development," in Elliot W. Eisner (Ed.), *Confronting Curriculum Reform*. Boston: Little, Brown, 1971, 17-49.

Bond, Guy L., and Robert Dykstra. "The Cooperative Research Program in First Grade Reading Instruction," *Reading Research Quarterly*, 2 (1967), 5-142.

Buswell, Guy T. "An Experimental Study of the Eye-Voice Span in Reading," *Supplementary Educational Monographs*, No. 17, University of Chicago Press, 1920.

Carroll, John B. "A Model of School Learning," *Teachers College Record*, 64 (1963), 723-733.

Cattell, James McKeen. "The Time it Takes to See and Name Objects," *Mind*, 11 (1886), 63-65.

Chall, Jeanne. *Learning to Read: The Great Debate*. New York: McGraw-Hill, 1967.

Chall, Jeanne. *Reading and Development*, Keynote Address, Twentieth Annual Convention. Newark, Delaware: International Reading Association, 1976.

Davidson, Helen P. "An Experimental Study of Bright, Average, and Dull Children at the Four Year Mental Level," *Genetic Psychology Monographs*, 9 (1931), 119-289.

Durkin, Dolores. *Children Who Read Early*. New York: Columbia University, Teachers College Press, 1966.

Durkin, Dolores. "When Should Children Begin to Read?" in Helen M. Robinson (Ed.), *Innovation and Change in Reading Instruction*, the Sixty-Seventh Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago, 1968, 30-71.

Ellson, D. G., P. Harris, and L. Barber. "A Full Test of Programed and Directed Tutoring," *Reading Research Quarterly*, 3 (1968), 307-367.

Gates, Arthur I. "An Experimental and Statistical Study of Reading and Reading Tests," *Journal of Educational Psychology*, 12 (1921), 303-314.

Gates, Arthur I. "A Study of the Role of Visual Perception, Intelligence, and Certain Associative Processes in Reading and Spelling," *Journal of Educational Psychology*, 17 (1926), 433-445.

Gates, Arthur I. *The Improvement of Reading*. New York: Macmillan, 1927.

Gates, Arthur I. *New Methods in Primary Reading*. New York: Bureau of Publications, Teachers College, Columbia University, 1928.

Gates, Arthur I., and Guy L. Bond. "Reading Readiness: A Study of Factors Determining Success and Failure in Beginning Reading," *Teachers College Record*, 37 (1936), 679-685.

Gates, Arthur, G. L. Bond, and D. H. Russell. *Method of Determining Reading Readiness*. New York: Bureau of Publications, Teachers College, Columbia University, 1939.

Gilbert, Luther C. "The Effect on Silent Reading of Attempting to Follow Oral Reading," *Elementary School Journal*, 40 (1940), 614-621.

Goodman, Kenneth S. "Dialect Barriers to Reading Comprehension," *Elementary English*, 42 (1965), 853-860.

Goodman, Kenneth S., and Catherine Buck. "Dialect Barriers to Reading Comprehension Revisited," *Reading Teacher*, 27 (1973), 6-12.

Gray, William S. *The Teaching of Reading and Writing*. An International Survey, Unesco. Glenview, Illinois: Scott, Foresman, 1956.

Gray, William S., and Bernice Leary. *What Makes a Book Readable*? Chicago: University of Chicago Press, 1935.

Holmes, Jack A. Research proposal submitted to and funded by the Carnegie Corporation, New York, 1959.

Holmes, Jack A. "When Should and Could Johnny Learn to Read?" in J. Allen Figurel (Ed.), *Challenge and Experiment in Reading*, Proceedings of the International Reading Association, 7, 1962. New York: Scholastic Magazines, 237-241.

Judd, Charles H., and Guy T. Buswell. *Silent Reading: A Study of the Various Types*.

Supplementary Educational Monographs, No. 23, University of Chicago Press, 1922.
Klare, George R. "Assessing Readability," *Reading Research Quarterly*, 10 (1974-1975), 62-102.
LaBerge, David, and S. Jay Samuels. "Toward a Theory of Automatic Information Processing in Reading," in H. Singer and R.B. Ruddell (Eds.), *Theoretical Models and Processes of Reading*, Second Edition. Newark, Delaware: International Reading Association, 1976.
Law, Alexander. *Student Achievement in California Schools: 1974-1975*, Annual Report, California Assessment Program. Sacramento: California State Department of Education, 1975.
Lucas, Marilyn S., and Harry Singer. "Dialect in Relation to Oral Reading Achievement: Recoding, Encoding, or Merely a Code?" *Journal of Reading Behavior*, 7 (1975), 137-148.
McCullough, Constance M. "Responses of Elementary School Children to Common Types of Reading Comprehension Questions," *Journal of Educational Research*, 51 (1967), 65-70.
McDermott, R. P. "Achieving School Failure: An Anthropological Approach to Illiteracy and Social Stratification," in H. Singer and R. B. Ruddell (Eds.), *Theoretical Models and Processes of Reading*, Second Edition. Newark, Delaware: International Reading Association, 1976, 389-428.
McNeil, John D., and Evan R. Keislar. Cooperative Research Project No. 1413, U.S. Office of Education. Reported by John D. McNeil, "Programed Instruction Versus Usual Classroom Procedures in Teaching Boys to Read," *American Educational Research Journal*, 1 (1964), 113-119.
Moore, Omar K. In Helen Rowan, "'Tis Time He Should Begin to Read," *Carnegie Corporation of New York Quarterly*, 9 (1961), 1-3.
Morphett, Mabel, and Carlton Washburne. "When Should Children Begin to Read?" *Elementary School Journal*, 31 (1931), 496-503.
Moynihan, Daniel P. "Sources of Resistance to the Coleman Report," *Harvard Educational Review*, 38 (1968), 23-36.
Otto, Wayne. "Developing a Skill Based Approach to Reading Comprehension," Proceedings of the National Reading Conference, 1976 (in press).
Preston, Ralph C. "Reading Achievement of German and American Children," *School and Society*, 90 (1962), 350-354.
Ruddell, Robert B. "A Longitudinal Study of Four Programs of Reading Instruction Varying in Emphasis on Regularity of Grapheme-Phoneme Correspondences and Language Structure on Reading Achievement in Grades Two and Three," Final Report, Project Nos. 3099 and 78085. Berkeley: University of California, 1968.
Russell, David. *Children's Thinking*. Boston: Ginn, 1956.
Russell, David G. "Reading Research that Makes a Difference," *Elementary English*, 38 (1961), 74-78.
Samuels, S. J. "The Technology of Reading: A Current View," Technological Report No. 2. University of Minnesota: Minnesota Reading Research Project, 1973.
Samuels, S. J. "Modes of Word Recognition," in H. Singer and R. Ruddell (Eds.), *Theoretical Models and Processes of Reading*, Second Edition. Newark, Delaware: International Reading Association, 1976.
Singer, Harry. "Substrata Factor Theory of Reading: Grade and Sex Differences in Reading at the Elementary School Level," in J. A. Figurel (Ed.), *Improvement of Reading through Classroom Practice*, Proceedings of the International Reading Association, 9, 1964. Newark, Delaware: International Reading Association, 313-320.
Singer, Harry. "Research that Should Have Made a Difference," *Elementary English*, 47 (1970), 27-34.

77

Singer, Harry, "IQ Is and Is Not Related to Reading," in S. Wanat (Ed.), *Issues in Evaluating Instruction*. Arlington, Virginia: Center for Applied Linguistics, 1977a, 43-63.

Singer, Harry, "The SEER Technique: A Noncomputational Procedure for Quickly Estimating Readability Level," *Journal of Reading Behavior*, 7 (1975), 255-267.

Singer, Harry. "Measurement of Early Reading Ability," in W.G. McCarthy (Guest Editor), *Contemporary Education*, 48 (1977b), 3, 145-150.

Singer, Harry. "Resolving Curricular Conflicts in the 1970s: Modifying the Hypothesis, It's the Teacher who Makes the Difference in Reading Achievement," *Language Arts*, 54 (1977c), 158-163.

Singer, Harry, and Sherrel Beasley. "Motivating a Disabled Reader," *Thirty-seventh Yearbook of the Claremont Reading Conference*. California: Claremont College, 1970, 141-160.

Singer, Harry and Dan Donlan. *Reading and Learning from Text*. In preparation, 1978.

Singer, Harry, and Alan Rhodes. "Problems, Prescriptions, and Possibilities in High School Reading," in W. Otto, C. Peters, and N. Peters (Eds.), *Perspective on Reading Problems*. New York: Allyn and Bacon, 1977.

Singer, Harry, S. J. Samuels, and Jean Spiroff. "Effect of Pictures and Context on Learning Sight Words, *Reading Research Quarterly*, 9 (1973-1974), 555-567.

Smith, Nila B. *American Reading Instruction*. Morristown, New Jersey: Silver Burdett, 1934.

Strang, Ruth. *Problems in the Improvement of Reading in High School and College*. Ephrata, Pennsylvania: Science Press, 1938.

Strang, Ruth. *Explorations in Reading Patterns*. Chicago: University of Chicago Press, 1942.

Terman, L. M., and Margaret Lima. *Children's Reading*. New York: Appleton, 1929.

Thorndike, Edward. "Reading as Reasoning: A Study of Mistakes in Paragraph Reading," *Journal of Educational Psychology*, 8 (1917), 323-332.

Many teachers would agree that flexibility—or the ability to adjust reading rate according to one's purpose for reading—is one of the hallmarks of a skilled reader. Surprisingly, the research literature indicates lack of flexibility even among the highly educated. Phyllis Miller points out reasons that these findings are erroneous. What are some reasons that these findings of lack of flexibility are wrong? What is wrong with trying to determine a person's typical reading rate? What are the proper testing procedures in determining whether a person is capable of adjusting reading rate? What instructional principles can be drawn from the concept of rate flexibility?

Considering flexibility of reading rate for assessment and development of efficient reading behavior

PHYLLIS A. MILLER
Augsburg College

Introduction

College and secondary level students are required to do both extensive and intensive reading. Many students are uncertain of their abilities to accomplish their reading tasks efficiently and request assessment and procedures for improvement in this area. Valid assessment and clear conceptions about rate relevant to instruction involve understanding and accounting for the factors that influence rate.

The concept of flexibility of reading rate describes those factors and offers an interpretation of the reading rate behavior of fluent readers. The flexibility concept has also provided useful predictions regarding rate behavior which certain research has confirmed.

Both the concept of flexibility and the recent studies supporting it imply principles applicable to assessment of rate and to instructions designed to modify it. The intent of this paper is to discuss those principles from the perspective of the conceptual and research base. A general description of current reading rate assessment instruments and instructional materials will be included to emphasize the need for much revision in this area.

Present assessment of rate

Presently available on the market are two general methods used to assess rate in standardized tests (Buros, 1972). Both methods are inadequate because they do not reflect a conceptualization of rate as it is viewed from the stance of reading flexibility.

One method consists of having the reader read a single easy passage for a specified time period. Directions for the tests using this method are evasive about the purpose for the reading and about the type of material contained in the passage. Following the reading of the passage, the reader answers a series of questions without referring back to the passage.

The other method requires the reader to use a proofreading approach for many easy sentences about varied topics. The reader either may have to cross out an inappropriate word or choose a correct response in a modified cloze procedure.

A single rate score is derived from the use of either method, although, in some tests, an accuracy score is also given for a particular rate.

Regardless of which method is employed in the assessment of rate, the information gained is limited and, even worse, misleading, relative to description and diagnosis of a person's level of reading efficiency. In addition, though of less consequence, the tests often do not contain material representative of that faced by students in reading situations at the secondary or college levels. Several changes must be made in rate tests in order for them to be valid measures of this area of reading behavior.

Principles applicable to assessment of rate implied by the concept of rate flexibility

Measure more than one aspect of rate

The first principle applicable to assessment is that more than a single aspect of rate must be measured in a test claiming to describe or diagnose a person's reading efficiency. The rationale for this is based upon the idea of reading flexibility which purports that an individual's reading rate is not a unitary phenomenon, rather, it can vary at one point in time in predictable ways. An individual does *not* have a single base rate.

Flexibility is the behavior, usually measured in terms of rate, by which a person either changes his reading speed or changes his reading strategies, or possibly both, depending upon the nature of three factors—the reader himself, the text, and a factor interacting between the reader and the text.

According to the notion of reading flexibility, a primary factor pertaining to the reader is his comprehension purpose, determined either internally or by external instructions. Reader purpose can refer to different things; for instance, it can refer to intellectual processes such as perceiving information literally, making inferences or critical judgments, or making affective responses. Or it can refer to kinds of information sought in a

passage, either general or detailed. The kind of process and type of information sought influence the strategy and rate with which a reader proceeds.

A textual factor is the difficulty of the discourse, usually determined by a non-personal measure such as a readability formula, or by the cloze procedure. Formulae generally account for sentence length and syllable variations or word frequencies. In actuality, difficulty is also dependent upon the reader's own developmental level at a certain point in time. Other textual factors influencing reading rate behavior include content type and style.

A factor interacting between reader and text is the amount of familiarity a reader may have with the content and form of the text. This is predicated upon the idea discussed in psycholinguistic approaches to reading (Smith, 1973) which claims that a reader draws upon his background store of nonvisual information when he processes text. This information consists of his past experiences with the world and with various language forms and symbols. During the reading process there is a trade-off between the visual and the nonvisual information. The more nonvisual information a reader has in store to draw upon for a particular piece of reading material, the less he has to rely upon the visual input in the text to gain information.

Variations in one or another of the three factors (reader purpose, textual difficulty or type, and familiarity) theoretically influence variations, in a predictable direction, of a fluent reader's rate. As Nacke (1971) has suggested, the flexibility concept is based on the assumption that a reader is capable of using a range of rates, in the sense of a continuum, i.e., from slower to faster. We may also assume that a fluent reader is capable of using various strategies, and that there is a relationship between type of strategy and rate.

In current standardized rate tests there is no way to determine which factor (reader, text, or familiarity interaction) is influencing the rate displayed by the reader for that test. Also, a single measurement, as is the common output of these tests, cannot account for possible variations in reader behavior (as indicated by rate) that occur as a result of variations in any of the three factors.

Certain research supports the idea that intraindividual reading rate changes as reader purpose, textual difficulty, or the familiarity interaction changes between reader and text.

Samuels and Dahl (1975) found that when different instructions were given regarding reading purpose, students read materials of the same difficulty levels at significantly different rates. The investigators asked fourth graders and college students to read two articles for two purposes—either general or detailed information. The readers did adjust their rate according to their purpose.

Miller (1976) found that community college students read materials of familiar and unfamiliar topics (familiarity was

determined by student ratings in pilot testing) at significantly different rates. Difficulty level, writing form, and general subject matter area (family-school life) were controlled; however, one set of materials related to a familiar scene while the other related to an unfamiliar one.

Further, students read materials of the same familiarity level at significantly different rates when the difficulty levels changed from fairly easy to difficult. A change in either element, familiarity or ease, by itself produced the adaptability in rate. Both in combination brought even more adaptability, but not significantly more.

Dixon (1951) found that familiarity with a subject field influenced the reading rate of college professors and graduate students from three different university departments. The subjects read passages in their own subject area at significantly higher rates than they read materials in the other areas.

Thus, certain studies support the idea that there are several factors which influence rate and that a manipulation of any of them predictably brings about a change in a reader's rate. A reader slows down 1) to gain details, 2) for unfamiliar content, and 3) for text of higher difficulty. We may assume that an in-depth reading strategy through a text is slower than either a straightforward or a skimming strategy. To be valid, a test claiming to assess rate must measure more than one aspect of it.

Manipulate both reader and textual factors

The question becomes, which factors of reader, text, and interaction ought to be varied in an assessment instrument intended for description of a reader's rate behavior or for diagnosis of weaknesses in his reading efficiency? The answer is that a valid test of rate should manipulate both a reader purpose factor and a textual factor, and control the interacting factor of familiarity. Thus far, there is no published test which systematically manipulates these two factors. One published test (Raygor, 1970) does manipulate a difficulty factor.

If a student displays flexibility in rate with differences in a textual factor (such as difficulty-level change) it is still not certain that he will display flexibility for different reading purposes unless that is also assessed. Likewise, reading for various purposes must be assessed at different levels of difficulty in order to gain some reflection of the reader's ability to differentially process varying levels of text.

There has been an attempt in some research to compare the relative effect of these two factors upon rate. Letson (1959) attempted such a comparison, but his study had many weaknesses,

so that his conclusion of more influence upon rate variation of difficulty than of purpose is subject to question.

Levin (1967) carried out a more complex study, aimed in part at the same kind of comparison. Unlike Letson's findings, she reported that there was little difference between flexibility according to purpose and flexibility according to difficulty. (Interestingly, there was some tendency for the better readers to vary their rate more for purpose, and for the poorer readers to vary their rate more for difficulty.)

Control the familiarity factor

Familiarity of topic is a variable that need not be manipulated in a test; however, the necessity to control for its effect is very important when assessing rate. If it is not controlled, it can have a confounding effect upon a rate score. This is particularly true when assessing rate between levels of difficulty.

For instance, if a student is familiar with the passage topic which is easy in terms of readability level, but unfamiliar with the passage topic that is difficult, more weight may be given to the effect of difficulty levels than is warranted. Conversely, if a student is more familiar with the content topic of a difficult selection than with that of an easy selection, he may display little flexibility between the two passages.

Miller (1976) looked at the interaction effects upon rate of the factors of purpose, familiarity, and difficulty. She found that a greater effect of familiarity level occurred for easy passages than for difficult ones. That is, students slowed their rate more from familiar to unfamiliar easy passages than they did from familiar to unfamiliar difficult passages. This same pattern occurred for students reading for either general or detailed purposes. There was no difference in comprehension scores between levels of difficulty or levels of familiarity.

This is further implication that interpretation of scores is subject to question for existing rate tests which use an easy passage or easy sentences with no topic familiarity control.

Make provision for reader set

A procedure to provide adequately for reader set is an important element to consider in the development of rate assessment instruments. Set refers to the adoption of a reading strategy or rate based upon the anticipation of a certain kind of question to be answered following the reading and upon the general readability characteristics of the text. When a reading purpose is externally imposed upon a reader (as in a testing situation), the internal acceptance by him of that purpose can present a procedural problem.

Until a few years ago, much of the reported research concerning flexibility indicated that little rate flexibility existed in the reading behavior of fluent readers. Those studies, failing to provide for reader set, led to misleading conclusions about reading flexibility in general.

Many of those studies investigated the effect of purpose variation upon rate (Boyd, 1966; Hill, 1964; Hurculane, 1961; McDonald, 1963; Spache, 1956; Theophemia, 1962). Usually, the readers were given several reading passages and told to read each passage for a different purpose. In the instrument reported on by Spache there was one passage, and students were told to read it three times for three purposes. A generalization arising from this group of studies was that most students were not flexible in their reading behavior.

However, a problem basic to all of those investigations, including those reporting on test instruments offered on the market, was that there was no provision to insure that subjects really understood or appropriated the purposes for which they were reading. In addition to the lack of adequate overall procedure were specific weaknesses in the directions, which included describing purposes vaguely or being inconsistent with the reading strategy or rate suggested for use with the purpose to be achieved.

The difference in findings concerning flexibility of more recent studies from those of the foregoing studies appears to relate to the degree to which the readers had an opportunity to establish a set appropriate for the particular purpose required by the reading task.

McConkie, Rayner, and Wilson (1973) found significant differences in rates with groups of college students for a particular passage followed by different types of questions. They used six passages of equal difficulty and wrote five types of questions for each passage. They compared the reading time on the first and last passages for subjects given one or another of the question types. On the first passage there were no differences in rate, but there were significant differences in rate on the last passage. They found that students could adjust their speed according to the type of questions they expected.

The Samuels and Dahl study (1975) was carried out with a twofold intent: first, to test the hypothesis that readers do adjust their rate according to their purpose; and second, to establish test conditions for display of this flexibility. Students read two stories or articles for two purposes; one was read to get a general overview, and the other was read to answer detailed questions. Each test was preceded by a practice test passage and set of questions. The practice was to indicate to the reader the type of questions he would have to answer in order to have a considerable degree of confidence in the directions.

This study demonstrated the importance of establishing

unambiguous test conditions when expecting subjects to read in different ways. Like McConkie *et al*, Samuels and Dahl found that rate is influenced by the kind of test a reader anticipates.

It is clear that rates of reading for the same passage can be considerably different when different sets are established for the reading task. A workable method to account for set is to use an adequate sample test preceding each different purpose, along with clearly specified purposes.

There are certain other aspects of reader set which also must be considered in rate tests. These pertain to difficulty level variations. Rothkopf and Coatney (1974) concluded that rate was dependent on two aspects of readability of text; the first was the readability level of currently-read material, and the second was that of recently-read materials. They found that subjects who first read ten paragraphs of difficult material read succeeding paragraphs of standard material more slowly. This effect was not as strong when standard level material was followed by difficult level material.

Set-like effects regarding difficulty imply two requirements for test construction in addition to sufficient contrasts in difficulty levels. Easier material should be read first, and passages of sufficient length must be used to offset the prevailing influence of set brought about from reading material of another difficulty level.

Current instructional materials for rate improvement

Certain principles applicable to instruction regarding reading efficiency emerge from the concept of reading rate flexibility. However, these principles appear to be absent in many of the current materials containing information and/or practice for improvement of rate. Unfortunately, the type and effectiveness of instruction in many programs designed for the development of reading efficiency reflect the kinds of materials available.

Generally, many materials contain misdirected emphases and useless exercises. The kinds of speed and perceptual improvement exercises often offered lack conceptual and research validated rationale.

For example, there are many single difficulty level (easy) kinds of exercises; little evidence exists that practice reading them has transfer effect to other difficulty level materials. Many short drills in reading speed abound; seldom are reading assignments that students face equally short.

There seems to be no rationale for the type or inclusion of many of the questions following exercises in practice books, except that an exercise must contain a sufficient number of these questions. Purpose for reading is not established, except that a reader is told to try to go faster. Reading, as a behavior, seems to

exist in a vacuum, rather than as an avenue for learning or enjoyment.

According to a content analysis of a sample of college reading manuals by Bahe (1970), over 50 percent of the manuals contained exercises for eye-span development, columnar phrasing, and digit-word recognition. Though there is little evidence regarding the effectiveness of these exercises for developing reading efficiency, the manuals containing them still are found in reading centers.

Instructional materials and techniques that are intended to develop efficiency ought to reflect principles consistent with the idea of reading flexibilty.

Principles applicable to instruction implied by the concept of flexibility

Develop awareness of time, strategy, and level of effectiveness for reading tasks

There appears to be an implicit assumption in materials available for instruction and practice in rate and comprehension that if students are to read something well, "for comprehension," they should not be timed. Further, if they are to read something superficially, "for rate," then they may be timed. This notion is missing an important point. That is, when moving toward the goal of greater efficiency, an awareness at some level of time, strategy and degree of effectiveness is necessary for whatever the task. Each element has its value in the overall process and must be integrated in the instruction.

It is easy to promote an awareness of time in brief exercises timed for speed, but not as easy to attend to that with longer, more difficult assignments emphasizing content. As a result, a perspective of developing overall efficiency evaporates.

However, for instructional applicability, timing need not be always in words per minute, or some such relatively careful measurement as in assessment instruments or published classroom exercises. Gross estimates can be made, for efficiency comparison, of time to read chapters or even books, with various levels of understanding, when applying different reading strategies and speeds.

A perspective of flexibility implies that a student should know specifically, before reading, what his comprehension purpose is, either within or between materials. He should be able to apply strategies to accomplish that purpose in a way that he may evaluate his use of those strategies, in terms of both time and level of

accomplishment, for that task. This applies to assignments of any length, but may be particularly relevant to longer ones.

Do not overdichotomize reading and text study

The idea of flexibility implies that it is not necessary to separate reading practice from certain forms of text study. Rates of straightforward reading strategies and in-depth reading or skimming strategies may fall at different points on a continuum, but they still exist as part of the same continuum.

As a result of too much emphasis on the dichotomy of these behaviors, students learn to take a relatively more passive stance when reading text versus a more active stance when organizing it in study. This results in inefficiency, as well as possible ineffectiveness for many reading situations. Helping a reader differentiate between purposes, and ways of dealing with those purposes, for varying texts enables him to be more active throughout the entire process.

Differentiate between purposes and strategies for reading

In instructing for a certain comprehension purpose with awareness of the time element, process directions must be given; i.e., various strategies must be described as well as the purpose given. A rationale should be given for using a particular strategy for either the carrying out of a certain intellectual process or the seeking of a certain kind of information from the discourse.

Besides providing purpose, strategy, and rationale, an instructor should provide sample questions which fit that particular purpose and which can be answered using certain strategies. In fact, various strategies for carrying out a purpose can be tried, timed, and evaluated to determine efficiency of each one.

For example, students can be given the purpose of seeking general points of information from a passage and thus be encouraged to use a skimming strategy. They can be given sample questions about points to be sought. They should be given a rationale for the skimming, such as the idea that gaining general points quickly can help provide a cognitive map for subsequent in-depth comprehension facility.

Then instructions for applying different skimming and nonskimming strategies can be given and these strategies tried and evaluated for both effectiveness and efficiency. For instance, a key word skimming technique, in which a reader attempts to look mainly at most content-bearing words throughout the paragraphs can be compared for effectiveness with another skimming strategy. That other strategy might be a form of the introduction, summary, and first sentence skimming technique, in which a reader reads only those portions of a text. Finally, straightforward reading can be

compared with both skimming strategies. Thus, a student can try a strategy, and then evaluate it, relative both to its usefulness for accomplishing his purpose and to the time it takes him to carry it out.

Instructors can make questions for a similar purpose for a series of passages, in order to provide students with practice in adopting various strategies for reading while being timed. Currently, most materials available on the market designed for efficiency improvement generally include a mixture of question types for a particular passage; thus, vague or confused reader purposes accompany rate practice. Students cannot be expected to know how to set their own differing purposes for reading, or focus upon certain kinds of information in passages, when they are questioned each time on a variety of unspecified purposes.

Use a range of difficulty levels for reading practice

While purpose can be held constant for a number of passages, a range of difficulty levels of materials should be used in that practice. As it is, often only easy material is used when students are expected to read for general impressions, and more difficult material is used when they are expected to read for details. This does not reflect the variety of uncontrolled difficulty levels to be faced by students in other reading situations.

Differentiate among content patterns in materials

Besides the usefulness of differentiating among purposes and strategies, it is also important to differentiate among the textual variations of content patterns and style. For instance, the idea development in narrative material often proceeds quite differently from that in expository material. A theme may emerge in subtle form from the plot or characterization in imaginative literature, while a thesis may be stated explicitly in exposition. Or a process organization in scientific material may proceed quite differently from an argument development in social science material. Thus, students should learn the patterns of differing kinds of textual discourse as well as a rationale and methods for setting and accomplishing different purposes within them.

Summary

The concept of reading flexibility maintains that a person's reading rate varies in predictable ways depending upon factors of text, reader purpose, and the familiarity of the reader with the text. Principles applicable to assessment and instruction regarding rate are implied by the concept and by the studies which support it.

To be valid, instruments for assessing reading rate must contain several features now lacking in current standardized rate tests. More than one aspect of rate must be measured; these measures should include both a reader purpose factor and a textual factor for adequate description or diagnosis of a reader's rate. Familiarity of topic must be controlled to lessen confounding influences. Provision must be made for the effects of set upon the rate.

There are several principles applicable to the development of reading efficiency which are implied from the idea of rate flexibility. However, they are not reflected in many current commercial materials.

There needs to be an awareness of time, strategy, and level of effectiveness regarding a reading situation. Straightforward reading and in-depth reading for text study should not be overdichotomized. Purposes and strategies for reading should be differentiated and possess a rationale for application. A variety of difficulty levels must be used for reading practice. Other varying textual features, such as content patterns, must be taken into account.

Sound measures of assessment and better instructional practices can result from consideration of these principles and the flexibility concept upon which they are based. These principles do not suggest quick, easy revisions of tests and materials; rather, they imply that a complexity exists in the description and development of reading rate efficiency. However, with the present urge to promote higher, as well as basic, levels of literacy, there is a strong need to work towards and accomplish these revisions.

References

Bahe, V. "A Content Analysis of Current College Reading Manuals," in G. Schick and M. May (Eds.), *Process and Pedagogy in Reading*, Nineteenth Yearbook of the National Reading Conference. Milwaukee: National Reading Conference, 1970, 11-25.

Boyd, R. "Rate of Comprehension in Reading among Sixth Form Pupils in New Zealand Schools," *Reading Teacher*, 20 (1966), 237-241.

Buros, O.K. (Ed.). *Seventh Mental Measurements Yearbook*. Highland Park, New Jersey: Gryphon Press, 1972.

Dixon, R. "Studies of the Eye Movements in Reading of University Professors and Graduate Students," in W. Morse, F. Ballantine, and R. Dixon (Eds.), *Studies in the Psychology of Reading*. Ann Arbor: University of Michigan Press, 1951.

Herculane, Sister M. "A Survey of the Flexibility of Reading Rates and Techniques According to Purpose," *Journal of Developmental Reading*, 4 (1961), 207-210.

Hill, W.R. "Influence of Direction upon Reading Flexibility of Advanced College Readers," in E. Thurston and L. Hafner (Eds.), *New Concepts in College—Adult Reading*. Thirteenth Yearbook of the National Reading Conference. Milwaukee: National Reading Conference, 1964, 119-124.

Letson, C. "The Relative Influence of Material and Purpose in Reading Rates," *Journal of Educational Research*, 52 (1959), 238-240.

Levin, B.J. "The Flexibility of Reading Rate," in J.A. Figurel (Ed.), *Forging Ahead in*

Reading, 1967 Proceedings, Volume 12, Part 1. Newark, Delaware: International Reading Association, 1968, 596-603.
McConkie, G.W., K. Rayner, and S. Wilson. "Experimental Manipulation of Reading Strategies," *Journal of Educational Psychology*, 65 (1973), 1-8.
McDonald, A.S. "Flexibility in Reading," in J.A. Figurel (Ed.), *Reading As an Intellectual Activity*, Proceedings of the International Reading Association, 8, 1963. New York: Scholastic Magazines, 81-84.
Miller, P.A. "The Effect Upon Reading Rate of Variations in Purpose, Familiarity, and Difficulty: An Investigation of Reading Flexibility at the Community College Level," unpublished doctoral dissertation, University of Minnesota, 1976.
Nacke, P.L. "Assessment of Flexible Efficient Reading," in F.P. Greene (Ed.), *Reading: The Right to Participate*, Twentieth Yearbook of the National Reading Conference. Milwaukee: National Reading Conference, 1971, 256-265.
Raygor, A. *Reading Test*. New York: McGraw-Hill, 1970.
Rothkopf, E. Z., and R. P. Coatney. "Effects of Readability of Context Passages on Subsequent Inspection Rates," *Journal of Applied Psychology*, 59 (1974), 679-682.
Samuels, S.J., and P.R. Dahl. "Establishing Appropriate Purpose for Reading and Its Effect on Flexibility of Reading Rate," *Journal of Educational Psychology*, 67 (1975), 38-43.
Smith, F. *Psycholinguistics and Reading*. New York: Holt, Rinehart and Winston, 1973.
Spache, G.D. "Diagnostic Tools," in O. Causey (Ed.), *Exploring the Goals of College Reading Programs*, Fifth Yearbook of the Southwest Reading Conference. Texas Christian University: National Reading Conference, 1956, 121-123.
Theophemia, Sister M. "Testing Flexibility in Reading," in J.A. Figurel (Ed.), *Challenge and Experiment in Reading*, Proceedings of the International Reading Association, 7, 1962. New York: Scholastic Magazines, 138-139.

A model is nothing more than a simplified representation of a complex phenomenon. One characteristic of a good model is that it can be used in decision making. Teachers make a number of decisions during reading instruction. Should all errors a child makes while reading aloud be corrected? Most would probably say no. But, then, how can one decide which errors should be of concern? Phonics activities may help children develop their decoding skills, but what kinds of activities are most likely to help students read the stories in their books? How can one approach instruction to help children understand their reading?

David Pearson suggests answers to these and other practical concerns. He does so through use of a psycholinguistic model of reading—a view which involves the reader's knowledge of meanings, grammar, and phonics. The model is explained in comprehensible terms, supported by research evidence, and applied in meaningful ways.

Some practical applications of a psycholinguistic model of reading

P. DAVID PEARSON
University of Minnesota

In recent years, several writers (Goodman, 1967; Smith, 1971; Samuels, 1970) have discussed language based models of the reading process for the purpose of providing those of us who deal with children's reading programs with a richer understanding of how children read and learn to read. Such models are often criticized because, it is argued, they provide little help for the teacher in the day-to-day process of making instructional decisions. In this article I want to demonstrate, to the contrary, that nothing is more practical to the teacher than a workable model of the process of reading and learning to read.

Review of a Language Based Model of the Reading Process

Receivers of language input (listeners or readers) use a variety of sources of information as they read or listen. Most prominent and ubiquitous for the reader are semantic-associational information, syntactic information, and grapho-phonemic (symbol sound) information.

Semantic-associational information includes our knowledge of what words refer to in the real world and how words are hierarchically related to one another.[1] But it also includes our knowledge of the fact that certain sets of words appearing in the same context "fit" together nicely, while other sets of words

Adapted from P. David Pearson, "A Psycholinguistic Model of Reading," *Language Arts*, 53, 3 (March 1976), 309-314. Reprinted by permission of the National Council of Teachers of English.

appearing in the same context are surprising to us. For example, the meaning of the terms *whale* and *horse* represent semantic information, as does the knowledge that both are mammals and hence related to man in some way. The associational aspect of this category of information is illustrated by our smugness when we find *whale* occurring in the same context with the terms *water, boat,* and *harpoon* and our surprise when it occurs in the same context with *farmer, wheat,* and *flax.*

Syntactic information refers to the ordering relationships among words in sentences. Even six-year-olds possess a great deal of information about how words are ordered within sentences to represent grammatically acceptable statements. In the sentences that follow, for example, they know that (1) and (2) but not (3) and (4) "make sense." Further they know that (1) and (2) are semantically equivalent to one another, that is, they mean the same thing.

1. The boy kicked the ball.
2. The ball was kicked by the boy.
3. ball boy the the kicked
4. kicked the the boy ball

While syntax and semantics are surely parts of what we refer to as context, syntactic information is, in my view, less useful as an aid to word identification and understanding *while we read* than is semantic-associational information. It is more useful for readers to know that the paragraph is about farm animals than it is for them to know that "the big..." precedes the word *horse* if their task is to decode *horse.*

On the other hand, syntactic information is critical when we speak of comprehension as product--or demonstrating to someone else that we have understood what we read. Given the semantically meaningless statement (5), we can (as can a first grader) answer questions (6), (7), and (8), solely on the basis of our knowledge of syntactic relations.

5. The roogle raznacked the sleefary because the sleefary had afranned his abelaide.
6. Who raznacked the sleefary...?
7. What was it that the sleefary had arfranned?
8. Why did the roogle raznack the sleefary?

Further note that (8) is a *why* type question, one usually thought of as requiring a higher order of understanding than *what* or *who* questions. As a matter of fact, most so called literal comprehension questions can be answered by referring only to the syntactic relations in a sentence. All one has to do is adopt a rule stating: "1) Read the question. 2) Scan the passage for the sentence containing the most words in common with the question. 3) Match up word for word between question and sentence. 4) Use the remainder as your answer." This "search and destroy" tactic is a very effective tool for students who want to demonstrate to us that they have understood

what they have read. Perhaps most of us recall using it from time to time, especially when the question was, "What are the major products of X?" We simply scanned the text until we found a sentence beginning, "The major products of X are...".

Grapho-phonemic (phonic, symbol-sound or, in Smith's term, visual) information is what novice readers learn in the phonics component of their reading program. It includes the knowledge that *b* translates as *buh* or that *b* is the first sound you hear when you say *bird* . It also includes the reader's knowledge, explicit or implicit, of phonics generalizatons, silent letters and letter-sound correspondence.

An important distinction exists between syntactic and semantic information on the one hand, and phonic information on the other. Both semantic and syntactic information form a part of data base (along with *phonological*—or speech sound—information) used by listeners to understand spoken language. Phonic information is peculiar to reading. Hence, the natural tie between oral and written language is through the base of meaning present in children's store of syntactic and semantic relations.

The model at work

The psycholinguistic model of reading proposed here suggests that real reading occurs when all three kinds of information are used in concert (see Figure 1). Efficient readers maximize their reliance on syntactic and semantic information in order to minimize the amount of *print to speech* processing (call this decoding, phonic, or grapho-phonemic analysis) they have to do. They literally predict what is coming and get enough grapho-phonemic information to verify their predictions. A single letter of a single syllable may be enough information to verify their predictions. For example, it doesn't take much visual or grapho-phonemic information to confirm the hypothesis that *telescope* fits into the sentence, "The astronomer looked through the _____ ."

Readers must vary the amount of attention they pay to the graphic information according to their familiarity with the content. This helps to explain why one can read *Time* magazine much more rapidly than a philosophical treatise. In terms of the model, *familiarity* with text can be translated into the degree of congruence between the syntactic/semantic information in the text and the syntactic/semantic relations stored in readers' heads. In simpler terms, familiarity implies *knowing* more about what is in a text.

Novice or poor readers are so bound up in their search for phonic information that they have little chance to attend to meaning. Good readers, on the other hand, because they attend to meaning, may often make unimportant errors in decoding words accurately. [2]

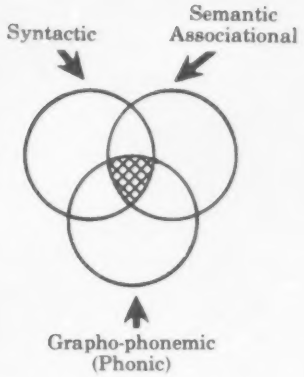


Figure 1. Sources of Information Used in Reading.
Mature reading occurs when all three sources of information are used in concert.

In all of these examples of the model at work, a common thread appears: the highly active role of the store of semantic and syntactic information already existing in readers' knowledge. Such a model is to be contrasted with views of reading which, even in the case of efficient readers, assign primary activity to *decoding* processes and relegate the semantic and syntactic processes to the role of verifying the phonological message. The present model will allow for situations (such as the novice or skilled reader confronted with unfamiliar text) in which decoding processes become the primary target of the reader's attention. But whenever possible, the model demands that the reader's attention focus on semantic/syntactic (meaning) processes.

Research support for the model

While there has not been a great deal of research that could be called a direct test of the proposed model, there have been a number of studies which provide partial support for aspects of the model. In addition, recent work in cognitive psychology (see Bransford and Franks, 1971; or a review of much of this research in Pezdek, 1975) has supported the general class of models of human memory (and comprehension) which emphasize the active constructive role of memory in modifying and schematizing new information processed by the human brain. The present model, with its emphasis on the active role of preexisting syntactic and semantic structures in determining reading speed, accuracy, and comprehension, is related to that general class of models.

The facilitative effect of context on word identification has been demonstrated in several studies. Goodman (1965) found that six-to-eight-year-old children reading in context correctly read 60-80 percent of the words they had previously missed in isolated word lists. Samuels (1970) found that semantic prompts prior to exposure of a target word facilitated speed of recognition. Tulving and Gold (1963) found an inverse relationship between contextual information and stimulus information: The amount of time needed to recognize a tachistoscopically presented word was an inverse function of the number of words in a preceding context. Pearson and Studt's data (1975) supported the notion that, for familiar target words, the contextual richness of a sentence context dramatically affected the amount of stimulus information six- and eight-year olds needed in order to identify the target word.

The relative influence of semantic versus syntactic constraints on comprehension and word identification has not been fully investigated. However, the cognitive studies of memory mentioned earlier as well as research by Pearson (1975) provides evidence for a view of comprehension which gives primacy to semantic rather than syntactic relations as basic units of cognitive storage.

In short, while the empirical evidence supporting the entire model is not overwhelming, validation for the model can be derived from studies supporting portions of the model as well as from studies which support the general class of models which the present model exemplifies.

Applications of the model

Use number one: To categorize language arts instructional activities. Virtually any language arts/reading activity can be classified as emphasizing one or more types of information available to the reader. One may ask of what use is such a categorization. My answer to that is that it gives you, as a teacher, a framework for knowing what you're about. Table 1 classifies a number of tasks according to the type(s) of information each emphasizes.

Use number two: To evaluate instructional activities. We can go beyond sheer classification to use the model for evaluation. Typically, any developmental reading program is a combination of a *skills activities component* intertwined in a scope and sequence and a *storyreading component.* One of the functions of a story component is to allow for a transfer of skills activities into a real reading situation. For example, in a linguistic series, after children learn the words *pad, fad, lad,* and *mad* in isolated vocabulary instruction, they get to practice reading them in sentences like "The mad lad can pad a fad." Likewise in a phonics series, after being

Table 1
Type of Information Emphasized by Various Language Arts Activities*

	A Phonic	*B* Semantic	*C* Syntactic
Emphasis on one type of information	**A** A worksheet with the headings *Rat* *Rate* and a number of words to be categorized, depending on vowel sound, under each word.	**B** Children are asked to classify 36 words into 3 categories: *Games* *Things to Eat* *Animal Names*	**C** Children are asked to choose a paraphrase of a standard sentence where only syntax has been changed: S. The cow ate the hay. 1. The hay ate the cow. 2. The hay cow the ate. 3. The hay was eaten by the cow.
Emphasis on two types of information	**AB** A worksheet with several problems which ask the children to use a picture cue to choose the correct word. ____ can ____ cane	**BC** Children are asked to listen to a sentence. Then tell whether the next sentence means about the same thing: "The cat chased the rat." "A rodent was chased by a feline."	
Emphasis on all three types of information	**ABC** A series of sentences with several choice points at which the children select the appropriate word: "A boy and a (man, mane) were walking down the (rod, road) toward a (tree, tred)."		

*Emphasis is the key word. It doesn't mean that the activity doesn't require the other types of information, just that one is emphasized more than the others.

exposed to a lesson on the phonic element *or*, they read stories containing sentences like, "The organ for chewing corn is the mouth."

The problem with the system is that the transfer jump from isolated skill instruction to reading is too large. Consequently, some children do not see the relationship between what happens in a skills activity and what they have to do when they read.

How does the model help us? By implication, what is happening is that the skills activities are concentrating on *phonic information only*. Then the children are asked to use the knowledge they have acquired in conjunction with context—for example, semantic and syntactic information. *But we don't show them how to do that.* There is a critical intermediate step left out: practicing the phonics skill in conjunction with semantic and syntactic information; in short, we need to expose them to *phonics in context*.

One can go a step further and use a decision rule which says that we will value most highly those phonics skill activities which allow children to utilize the most semantic and syntactic information while they are "cracking the code." Conversely, we will value least highly those phonics skill activities which are most isolated from context. Referring to the examples in Table 1, the tasks, A, AB and ABC, would rank ABC-AB-A. But even if we allow *initial instruction* to occur in isolated contexts, we will always require the intermediate phonics-in-context step *prior to* the attempt to have children transfer the skill to a *real reading* situation.

Use number three: To analyze reading disability cases. In the past decade and a half, as our awareness of language and/or culturally unique children has increased, many of them have been funneled into remedial reading programs. Usually there has been a diagnosis that such children need remedial help in word recognition and comprehension, accompanied by a prescription for a heavy dose of phonics and pleasure reading in high-interest, low-vocabulary materials.

Many of those prescriptions were probably appropriate. But if they were, it was often accidental. The failure of diagnosticians to look at reading failure in the context of language has, in my opinion, been one of our most serious educational shortcomings. It has led to situations in which we treat a child whose basic problem is a language difference as if he had a reading deficiency.

The following profiles of skill achievement (see Figure 2) help to illustrate this point. Pupils A and B both exhibit poor comprehension. But the reasons for their poor comprehension are drastically different. Pupil A does not comprehend well because he/she literally cannot read the words. Pupil B, on the other hand, can read the words; he/she simply does not know what they mean. Assuming they both speak standard English, and that other factors, like motivation, are constant, we ought to prescribe very

different remedial programs for each of them. Pupil A may need a heavy dose of phonics. Pupil B doesn't really have a "reading" problem; what he/she really needs is a strong experience-oriented program in concept development. But it *needn't be based in reading*. It could be entirely a *listening-speaking-sensing* program. Our model, because it assigns such a central role to the language basis of reading, would not only allow for, but likely encourage, such a program for pupil B.

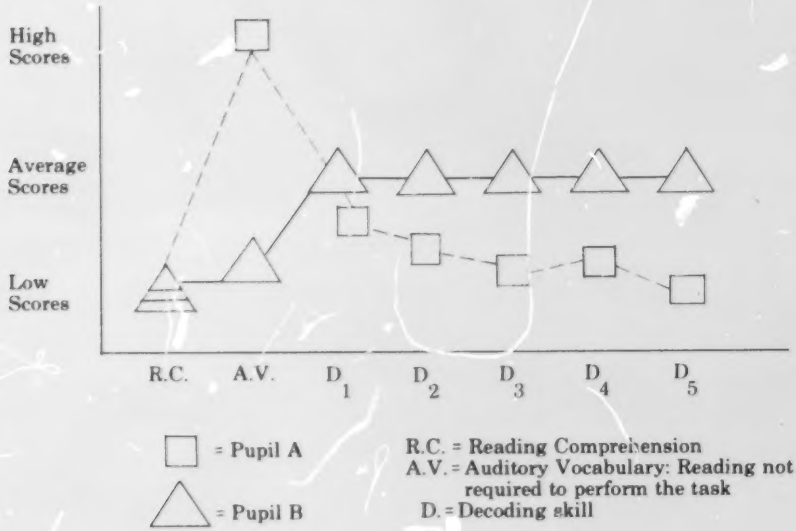


Figure 2. Sample Profiles.

Use number four: To study efforts made in oral reading. The conventional wisdom in looking at oral reading errors has insisted that an error is a thing to be corrected. This has lead to practices such as asking children to respond to a flashcard of the missed word fifteen times in order to remediate the error, *or* making lists of words missed and having the children practice reading them to one another. Such practices have a seductive rationale behind them; take the errors the children actually make in reading and ask them to practice them. What could be more relevant, more functional?

These practices fail to consider why the so-called error was made in the first place. They fail to force us to make inferences about what a child was thinking when he read the material. In short, they fail to encourage us to look at errors in the context of the language—both what was on the page and what was in the child's head—in which they occurred.

When one puts language into an analysis of errors, new perspectives arise. First, we will probably find a substitute for the word, *error*. We are likely, as the Goodmans have, to coin a term such as *miscue*. We do this because we realize that many times errors result from what children are thinking about when they read. In a sense, the language in children's minds miscues them, predisposes them, to respond differently from what is on the page. For example, if a child said /was/ for "saw" in the sentence, "Mary saw a girl," it may have made good sense to say /Mary was a girl/.

Additionally, we will be predisposed to regard certain kinds of errors, in certain contexts, as unimportant. Regressions, for example (starting a sentence, getting halfway through, and going back to start over), will be seen as good strategy. In most instances, children regress because they realize they have made an error.

Certain small word or function word confusions (for example, the/this, a/the, when/where, that/this) will often be seen as habitual miscalls that don't interfere with comprehension.

We will regard children who speak black dialect and say /poz/ for "supposed" or /be/ for "is" as thinking readers rather than blundering readers. For we realize they are translating the visual form "supposed" into the appropriate phonological representation in black dialect /poz/. That is a highly complex skill, one that does not deserve our disdain. We will make similar judgments about the Chicano children who say /boys/ for "voice" or /shursh/ for "church." Now you may want to work on teaching standard English (that's an entirely different issue), but at least you will avoid classifying a language process as a reading error.

We will not be overly concerned about the substitution of /giant/ for "gigantic" in the sentence, "What a gigantic cake," because it was a "good" error; that is, it made sense and wasn't that different from the word on the page.[3]

Other types of error patterns will be regarded as even more serious than before. I recall a boy in the very first elementary class I ever had. His strategy for oral reading was to say anything that had some visual similarity to the first part of the word he was attempting to decode, a practice which resulted in output like /The dover souping the car/ for "The driver stopped the car." The seriousness of error patterns such as this one stems from the child's total disregard for *language sense* while reading. In other words, one of the goals of good oral reading instruction should be to help children use their knowledge of oral language to monitor their oral reading output.

The child whose error pattern is illustrated had somehow learned that oral reading doesn't have to make sense in the same way as speech.

How does such an attitude become established? There are several possible contributing factors. First, if children are exposed to content foreign to their oral language repertoire, they may never see the relationship between speech and print (note, for example, the peculiar patterns in early linguistic readers and traditional basal readers). Second, the oral reading atmosphere may be so anxiety provoking that they learn it is better to suffer a little embarrassment by saying dumb things than it is to endure a lot of anxiety by trying hard to read what is on the page. Third, no one may ever have pointed out to them that oral and written language, while differing in form and precision, stem from a common source—their experiences.

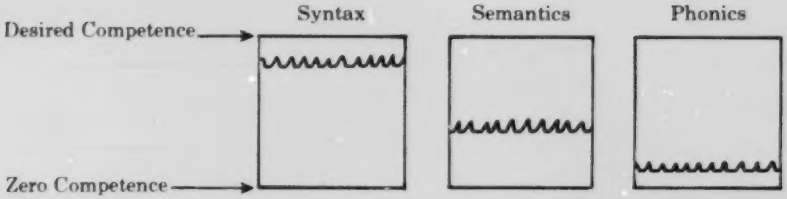
In terms of providing help to a student who has developed such an attitude toward reading, the model would suggest that initial instruction might occur in listening mode. One strategy is to purposely read to the children sentences or paragraphs containing anomolous words, asking them to stop you when they hear something that makes no sense. Another is to give them the correct text and have them circle the words on which you make errors when *you* read a specially prepared text.

Finally, when evaluating the seriousness of any error or error pattern, we will be likely to apply this criterion: Does the error interfere with the child's understanding of the story? A good way to test this is to ask the child, after finishing a selection, to tell you what he or she has read. Another strategy is to ask a question which requires an understanding of a miscued word. Many times a child will misread a word but, in retelling the story, say it accurately. Perhaps it was a slip of the tongue. Perhaps later context triggered a mental correction.

Use number five: Pinpointing similarities and differences among approaches to reading. Let us assume that all approaches to reading, regardless of differences in method, share a common goal: the development of proficient, comprehending readers. In terms of our model, such a goal implies that all such readers will achieve a proper integration of syntactic, semantic-associational, and phonic information. Differences among approaches center upon how each approach attempts to utilize (or ignore) certain sources of information in trying to bring the other source(s) of information into proper balance. The diagram in Figure 3 may help to make the point.

Relative to desired competence, the novice reader possesses a great deal of syntactic, a moderate level of semantic, and very little phonic information. Reading approaches differ in terms of the strategies used to move each level of information toward desired competence.

For example, contrast a linguistic approach with a language



= Estimate of the actual competence possessed by the novice reader.

Figure 3. A comparison of the relative proportion of desired competence possessed by a novice reader in the areas of syntactic, semantic, and phonic information.

experience approach. The language experience rationale is as follows: The child already possesses competence in syntactic and semantic information. Therefore, we should capitalize on those existing strengths to pull competence in phonics into line. The most sensible way to do this is to utilize the children's own language as the content of reading instruction. In this process, hopefully, the children will be able to "read" the content because of its familiarity to them and, in so doing, will abstract the appropriate phonics information from the corpus of phonics relations present in the words to which they are exposed. This can be thought of as a "bootstrapping" strategy.

A contrasting rationale underlies linguistic approaches: Since the children are relatively competent in syntax and semantics compared to phonics, what we need to do is to attack the weakness directly. Therefore, we teach phonics skills with little or no attention given to semantics or syntax.[4] In fact, if we rely on syntax or semantics to assist in the development of phonics skills, we may harm children's progress by deflecting them from the task at hand. Once phonics skills are brought up to par, they can be integrated with preexisting competence in semantics and syntax, and *voila*, the children read! This can be thought of as a "frontal assualt" strategy.

Other approaches to teaching reading probably fall somewhere in between these two cases. So called phonic or phonic-linguistic approaches are similar in spirit to so called linguistic approaches.[5] They emphasize phonics information often at the expense of contextual (syntactic and semantic) information. Traditional meaning-emphasis approaches tend toward the spirit of language experience in that phonics information is typically taught using a corpus of words already known by the children. It is probably true that they rely more on semantic rather than syntactic information in pursuit of phonics information; that is, the word rather than the sentence is the typical unit of analysis in meaning emphasis phonics lessons.

Use number six: To provide a blueprint for comprehension instruction. It has become almost trite to contend that the basis of concept development and prose comprehension is direct experience. Almost equally trite is the corollary that comprehension will be facilitated or debilitated according to the degree of prior experience with the concepts in question. Like so many other trite aphorisms and admonitions, the triteness of these statements probably stems from the fact that they are true.

Yet, reading comprehension does not depend solely on experience; for if it did, we would not be able to understand novel concepts or situations when we read about them.

An extension of the semantic-associational component of our model can help in understanding the relationship between prior real-world experience and comprehension of apparently new concepts or scenarios just as it can help in suggesting teaching strategies we can adopt to foster comprehension.

At the outset, in describing the semantic-associational componet, I indicated that it included not only our knowledge of word meanings but also our knowledge of the often hierarchial relations among meanings. Our knowledge is schematically rather than randomly organized. Analogously, with events, knowledge is schematically organized around typical scenarios or scripts (to borrow a word from the artificial intelligence discipline--Schank, 1975). These structures of knowledge are built from and are constantly modified by our experience. Thus, to understand new information means either to incorporate it into a preexisting schema or to modify a preexisting schema so that it can be incorporated.[6] To take a specific example, when young children who call both dogs and cats dogs are told that the Siamese they have just called a dog is a cat, they are forced to modify their dog schema and create a new schema for cats. Later on when they call a minx a cat or a collie a dog, they are understanding by incorporating that new information into existing schema.

But too often we have asked the wrong question—and held inappropriate expectations—regarding prior experience. Instead of asking, "What direct experience have the children actually had with tigers?" we should be asking, "What experience have the children actually had that I can use in helping them understand the concept of tigers?" The children may never have seen, touched, or smelled live tigers but they may have seen pictures or television programs. Lacking even that vicarious experience, the children may have had direct experience with domestic cats and may have seen wild predatory animals in a zoo. In the latter case, the appropriate channel of access to the concept of tiger may be through the class relationship that both domestic cats and tigers share (both are felines or cats in the generic sense) and/or through the common attribute of predation that tigers share with other beasts, that they are not feline (e.g.

wolves, jackals). Consider how prevalent such a strategy is among both adults and children when they try to explain something new to a friend. Quite often the explanation begins, "It's sort of like a . . . ." and continues, "but it's a little different in that . . . ." They try to get the new concept across by tying it into some experience with which they judge the friend to be familiar. These strategies follow quite naturally from a schematic view of the way knowledge is stored in memory, as is implied by our model.

In short, the crucial aspect of concept development and comprehension in vicariously-oriented school environments is not direct experience with the new information in question but rather how to access the new information through existing schema that may be based upon direct experiences.

Concluding statement

I have tried to convince the reader that a good model provides an excellent framework for making practical instructional decisions. Such an endeavor has been criticized from professionals at both ends of the theory/practice continuum.

Some theoreticians view the attempt to integrate theory and practice as a dilution (and possibly a misuse) of their theories. They believe that theories and theoreticians have no obligation to develop implications for classroom practice. A theory, they argue, should be judged on the grounds of how elegantly, parsimoniously, and precisely it explains the cognitive phenomena in question. Some practitioners argue that grand theories won't wash in the complex ecology of the classroom. They break down. They fail to speak to the exigencies of the teaching learning situation.

While I recognize the motivation for each of these types of criticism, I cannot accept them. Models of the process of reading and the process of learning to read are, in my opinion, invaluable for teachers. A good model ought to make predictions relevant to a variety of specific situations. Also, it should endure over time and between circumstances. In short, if we possess a good model of reading, we ought to be able to make adaptions to specific circumstances and specific children. I have tried to iterate some examples of particular adaptions, or applications, or a psycholinguistic model. I encourage readers to familiarize themselves with a more profound treatment of such a model and to generate their own practical applications.

Footnotes

1. Most writers speak only of semantic information. I add associational to the term to emphasize the fact that the body of information is related to experience and consequently to "sets" or expectations we have about what words are and are not congruent with one another.

2. The reader who wishes a more detailed and sophisticated explication of the model I have attempted to summarize should consult the Smith reference (1975) listed at the end of this article.

3. I don't want to push this argument to the point of saying that a child who says /alligators/ for "crocodiles" has made a "good error." While the two words may be interchangeable in many contexts, there would be several in which they weren't, e.g. "_____ live in Africa." Furthermore, that child is violating an important principle of English writing, i.e., the fact that it is alphabetical in nature—letters do represent sounds, at least to some degree.

4. I do not want to imply that linguistic approaches teach phonics correspondences or generalizations directly. They do not. Yet they do arrange and sequence lessons in such a way as to encourage the children to abstract phonic information without reliance on context.

5. I have tried, in vain to discover what makes a linguistic approach "linguistic." Since all approaches use language (eventually), they could all be regarded as linguistic. The only common element I can find among linguistic approaches is that they have a linguist on their authorship team. They might better be labelled "alphabetic" approaches since they tend to emphasize the alphabetic nature of writing.

6. In Piaget's terminology, these two alternatives would be called assimilation and accommodation respectively. In Frank Smith's terminology (1975) they would be called comprehension and learning in that order.

References

Bransford, J., and J. Franks. "The Abstraction of Linguistic Ideas," *Cognitive Psychology*, 2 (1971) 331-350.

Goodman, K. S. "A Linguistic Study of Cues and Miscues in Reading," *Elementary English*, 42 (1965), 639-643.

Goodman, K. S. "Reading: A Psycholinguistic Guessing Game," *Journal of the Reading Specialist*, 4 (May 1967), 125-135.

Pearson, P. D. "The Effects of Grammatical Complexity on Children's Comprehension, Recall, and Conception of Certain Semantic Relations," *Reading Research Quarterly*, 10 (1975), 155-192.

Pearson, P. D., and A. Studt. "Effects of Word Frequency and Contextual Richness on Children's Word Identification Abilities," *Journal of Educational Psychology*, 67 (1975), 89-95.

Pezdek, K. "Arguments for a Constructive Approach to Memory," unpublished manuscript, University of Massachusetts, 1975.

Samuels, S. J. "Recognition of Flashed Words by Children," *Child Development*, 41 (1970), 1089-1094.

Schank, R. "The Structure of Episodes in Memory," in D. Bobraw and A. Collins, (Eds.), *Representation and Understanding*. London: Academic Press, 1975.

Smith, F. *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read*. New York: Holt, Rinehart and Winston, 1971.

Smith, F. *Comprehension and Learning*. New York: Holt, Rinehart and Winston, 1975.

Tulving, E., and C. Gold. "Stimulus Information and Contextual Information as Determinants of Tachistoscopic Recognition of Words," *Journal of Experimental Psychology*, 66 (1963), 319-327.

Using ideas derived from Piaget's theory of intellectual development, Frank Murray presents a view of reading readiness which is quite different from views held by many learning theoreticians. The learning theoretician would tend to view reading as a complex activity comprised of numerous subordinate skills, perhaps arranged in a hierarchy from simple to more complex skills. With this view, readiness consists of having mastered all the subordinate skills necessary to begin a new level of learning. Dr. Murray suggests there is a different type of readiness reflected in the child's level of thinking and ability to understand the process of reading.

Does Murray suggest that the teacher wait for the match to occur between the child's cognitive structure and the demands of the reading task? What are some things the teacher can do to bring about cognitive readiness for reading?

Implications of Piaget's theory for reading instruction

Frank B. Murray
University of Delaware

Piaget's theory of the development of intellectual competence is, strictly speaking, neither a psychological nor an educational theory, but rather a theory which makes a contribution to the discipline of genetic epistemology. This discipline is the branch of philosophy that treats the development and structure of knowledge. Piaget's contribution to it is largely methodological since he has provided a way of resolving classical epistemological problems through the careful observation of how children construct knowledge. Consequently, the recommendations from the theory for educational practice are quite indirect and loose. In the main, the practical implications of the theory for education have already been made by others, and in fact what has really happened is that the theory has provided a theoretical justification for a number of long standing educational innovations.

Any instructional innovation—progressive education, Montessori, open education, or discovery learning—which makes a major provision for the self-initiated activity of the pupil can be justified in the theory. On the other hand, the overriding curricular recommendation from the theory is that the curriculum sequence should match or reflect the stage of cognitive development of the child. That the curriculum sequence should be based upon the psychology of the developing child is not a novel idea, but what makes it interesting in this case is that Piaget has been able to be specific about the order in which specific curricular concepts and competencies develop.

The recommendations for reading instruction are limited somewhat by Piaget's view of the relationship between the development of language and thought in which the former is dominated by the latter. The theory holds that cognitive structure develops from a set of basic innate sensorimotor reflexes that are expanded through a series of qualitatively different stages as a

result of the character of the child's activity. Along with the initial endowment of reflexes or schemes, the child is endowed with two complementary mechanisms which function either to preserve cognitive structure (assimilation) or to modify it (accommodation). An overriding mechanism (equilibration) balances these two and serves to produce the appropriate blend of rigidity and flexibility that characterizes intelligent behavior. The particular character of the action (e.g., overt or covert; reversible or nonreversible; static or fluid) determines the particular stage of cognitive structural development. Cognitive structures are forms for ideas that are not merely true, but those which are necessarily true. What is developing in the child is a system of thought in which certain ideas have to be true (e.g., if A=B and B=C, then A=C) and a general methodology with which the child can validate truth. Language, according to this theory, initially contributes very little to this structural development because to understand information presented in language the child must have *already* constructed a scheme or structure into which to assimilate the information. As language is only a referential or sign system, one needs to have the referent in mind before one can understand the meaning of the words or signs of the language system.

This, of course, is an ancient argument. In the *Phaedrus*, Plato cited a myth in which the god, Theuth, who had invented writing (along with number, calculation, geometry, dice, and checkers) argued that writing should be made available to every person as a memory aid and a guide to wisdom. The king, Thamus, countered,

> Theuth, my paragon of inventors, the discoverer of an art is not the best judge of the good or harm which will accrue to those who practice it. So it is in this case; you, who are the father of writing, have out of fondness for your offspring attributed to it quite the opposite of its real function. Those who acquire it will cease to exercise their memory and become forgetful; they will rely on writing to bring things to their remembrance by external signs instead of on their own internal resources. What you have discovered is a receipt for recollection, not for memory. And as for wisdom, your pupils will have the reputation for it without the reality: they will receive a quantity of information without proper instruction, and in consequence be thought very knowledgeable when they are for the most part quite ignorant. And because they are filled with the conceit of wisdom instead of real wisdom they will be a burden to society (*Phaedrus*, 275).

Some eight centuries later St. Augustine continued the general argument by claiming that learning could not take place through telling. In *De Magistro (The Teacher)* he wrote (somewhat ironically) that a word was learned from the thing cognized rather than the other way around.

> What I am above all trying to convince you of, if I can, is that we do not learn anything by means of signs called words . . . we learn the meaning of the word—that is the signification that is hidden in the sound—only after the reality itself which is signified has been recognized, rather than perceive that reality by means of such signification.

The educational implications of this position, the idea that meaning precedes language, are not limited simply to reading instruction. They constitute a general pedagogical recommendation that teaching by telling is inherently ineffective unless a provision has been made for the child to construct—independently of language—the meaning behind the word. The specific recommendation for reading instruction is that a smaller portion of the primary school curriculum could be devoted to reading than is presently the case, since reading is not critical for early cognitive development. This of course is a controversial recommendation but one that has been put forth by Furth (1970) in *Piaget for Teachers* in the following manner:

> The average five to nine year old child from any environment is unlikely, when busy with reading or writing, to engage his intellectual powers to any substantial degree . . . . A school that in the earliest grades focuses on reading cannot also focus on thinking. It must choose to foster one or the other. Historically it has chosen reading . . . the first job of our elementary schools today should be to strengthen the thinking foundation on which any particular learning is grounded . . . if early schooling aims to emphasize and purposefully nourish the thinking capacity of the child, it cannot look to performance in reading and writing as the immediate criterion of success . . . . My whole contention is that early reading does not have an intrinsic relation to intelligence and that its one-sided emphasis implies an underemphasis of intellectual development (Furth, 1970, 4-6).

Furth's argument is not that children cannot learn to read in the primary school or earlier but that the instructional effort needed to teach reading is too great and should be placed instead in activities which directly promote cognitive development. The argument can be countered to some extent by the fact that cognitive development of children at this general age proceeds quite independently of schooling (Goodnow and Bethon, 1966; Mermelstein and Shulman, 1967), and that the effects of many training procedures for cognitive structure on children of this age are equivocal (Brainerd and Allen, 1971; Beilin, 1971, and 1977).

Furth's argument also is not one against the eventual teaching of reading. The positive implications of Piaget's theory for the teaching of reading center upon the limitations cognitive structure may place upon decoding and comprehension. The general

hypothesis is that concrete operational thought may be a prerequisite for reading. It seems to be a sufficient condition for reading, but not a necessary one because preoperational children in fact *do* learn to read. Presumably in these cases the preoperational reading skill is supported by some competence system other than the one Piaget has described for intelligence.

By the time children are seven years old, they can respond to a number of problems more or less as an adult would respond to them. Before that time children tend to give quite surprising responses, such as that their brother has no brother (egocentrism); that a flattened clay ball has more clay, weighs more, and takes up more space than it did before it was flattened (conservation); that the water level in a tilted glass is parallel to the bottom of the glass (absolute space); that there are more girls in the class than children (class inclusion); that if John is older than Jim, and Jim is older than Bob, John may or may not be older than Bob (transitivity); that a person cannot be two things at the same time like an American and a Catholic (centration); or that a rearranged group of objects has more objects than it had before the objects were moved about (number conservation). These children under seven tend also to focus their attention on only one aspect of a situation, such as noticing that a milk bottle and juice glass differ only in height while ignoring the difference in width. Similarly they seem incapable of ordering a set of ten or so sticks according to their lengths (seriation) because, presumably, this would require thinking of the same stick as both longer than the stick to its right and shorter than the one to its left. As well, all kinds of sorting and classification tasks from simple sorting on one dimension to more complex two by two (or more) classifications, are incorrectly performed by these children. Moreover, when they are pressed for their reasons for responding as they do they resort finally to elaborate and preposterous justifications (justification at any price). All of these peculiarities in the young child's responses are characteristic of the preoperational stage of reasoning and could presumably provide obstacles to decoding and almost certainly place limitations upon comprehension. In the space of three or four years, however, the child's responses in these areas become very much like the adult's.

Piaget hypothesizes that the ability to give adult-like responses in these instances rests upon the development of a system of operations. It is a system of reversible internalized actions that allows the child to mentally reconstruct the flattened ball as the original round ball (i.e., to reverse mentally the flattening transformation), to see himself as his brother's brother, to view the same object from more than one perspective or as on more than one dimension, etc. This system of operations, furthermore, has the properties of a mathematical group (viz. closure, associativity, identity, and inversion) and is both Piaget's description and

explanation of the qualitative changes in the child's thinking that are found between five and ten years of age.

These operations are taken to operate distinctly upon different content areas so that there appears to be little transfer between the child's solutions to problems. Thus, the child typically conserves number (that the number of objects in a set is unaffected by their spatial arrangement) before he conserves weight or volume, or he may be able to classify animals before vehicles although the principles of classification and deduction are the same in each case. The operations are said to be content determined or bound to specific *concrete* situations (hence, concrete operations). That these operations function in various aspects of the reading task is a major implication of Piaget's theory for reading.

Virtually all task analyses of the reading skill (Gibson, 1968 or Carroll, 1970) acknowledge that the child must first know a language to read. For Piaget, this acquisition is made possible by the development of representation (the ability to respond to stimuli that are no longer tangibly present), which signals the transition between sensorimotor knowing and the next stage (preoperational knowing). It is also acknowledged that reading requires the competence to dissect spoken words into sounds; to recognize and discriminate letters and graphemes; to decode graphemes to phonemes; to process more than one letter at a time; to process in left to right and up to down order; to know that written words are analogous to spoken words; to recognize words from such cues as configuration of letters, parts of the word, word context, and probabilities; and to reason, think, make inferences about what was decoded. A main implication of Piaget's theory is that the competence to do all these things rests upon the development of concrete operations. The corollary of this implication is that reading readiness should be viewed from this perspective.

It would appear that the construction of the concept of number, which Piaget views as a fusion of the development of cardinality (that a numeral stands for a class of things) and ordinality (that a numeral has a position in a series—seriation), should be no less difficult a construction, for example, than the construction of the concept of letter, which has a kind of cardinality and ordinality of its own and a place in a rule governed system. Like conservation of number, which results from various spatial transformations on a set of objects, there can be conservation of letter under various form transformations of case, size, and font. The notion is that the construction of the concept of letter is as dependent upon concrete operativity as are the concepts of number, length, substance, etc. Conservation of word (across transformations from active to passive, for example) presumably represents similar, but also untested, relationships between reading and operativity.

In a similar way, it is possible to conceptualize the dissection of spoken words into sounds and words into letters as a class inclusion problem. The recognition that words have spoken and written forms can be seen as another decentration problem. Moreover, the all important letter-sound correspondences may be viewed as instances of bi-univocal (square matrix) or co-univocal (triangular matrix) multiplications of classes and relations of sounds and letters. Indeed, Grant (1972) found that operativity, and in particular performance on bi-univocal multiplication of classses, predicted reading achievement quite well and, in fact, as well as language development measures predicted it.

Without quantity conservation it should be impossible, for example, to comprehend the point of the sentence, "Because he was on a diet, the fat man asked the waitress to cut his pie into four rather than eight pieces." The understanding of prose passages, the reasoning and making of inferences that define reading comprehension, also define comprehension in general and seem critically dependent upon such competencies as transitivity and formal operativity.

The evidence for the relationship between operativity and reading is largely correlational or limited to studies in which it is shown that conservers read better than nonconservers, or that good readers are better conservers than poor readers of the same age and IQ (Waller, 1976). There seems to be a reliable, low positive correlation (about .40) between conservation and reading performance as measured by such standardized tests as the Metropolitan Reading Readiness Test, (Brekke, Williams, and Harlow, 1973; Heatherly, 1972; Rausher, 1970). Stanford Achievement Test (Kaufman and Kaufman, 1972) or individually designed reading tasks (Goldschmid and Bentler, 1968; DeVries, 1974; Stack and Murray, 1976). The correlation between reading and conservation is reduced somewhat, but usually not appreciably, by partialling our general intelligence; in these studies, conservation is often more highly correlated with reading than IQ is correlated with reading. Performance on other operativity tasks is also positively related to reading, such as class inclusion (DeVries, 1974; Smith, 1971), seriation (Scott, 1969; Kaufman and Kaufman, 1972), absolute space and horizontality (Lovell, Shapton, and Warren, 1964), generic identity conservation, and transitivity (DeVries, 1974). Moreover, Briggs and Elkind (1973, 1974) have found that early readers score significantly higher on conservation than normal age readers of the same IQ.

Elkind and his colleagues have further identified a number of perceptual decentration abilities (systematic exploration of the stimulus, flexibility with figure ground reversals, transcendence over the content of parts in perception of the whole, and mental transportation of stimuli from one context to another). These

decentration abilities and reading achievement share a common factor (Elkind, Horn, and Schneider, 1965) and average readers of the same IQ as slow readers perform better on these decentration tasks than slow readers (Elkind, Larson, and Van Dorminck, 1965). Moreover, training on the tasks seems to lead to an improvement in word recognition and form (Elkind and Deblinger, 1969).

At the heart of initial reading is the ability to process more than one letter at a time and later to process more than one word at a time. This ability may be related to what Elkind has called *schematization* or the ability to transcend the parts of one sort to see a whole of another; an example would be to recognize that a configuration of various fruits had the shape of an animal. More specifically, the ability has been investigated directly by Farnham-Diggory and Hermon (1968). They taught a group of conservers and nonconservers the meanings of various logographs for the concepts, *jump, block, over, teacher, walk,* and *around,* and found that each group learned the logographs equally well. However, when the logographs were put together in a series, such as *walk-around-teacher* or *jump-over-block,* and the children were directed to act out what the logographs said, the conservers performed at a significantly higher level (7.54 out of 8.00) than the nonconservers (2.31). The conservers seemed able to synthesize the elements into a meaningful whole which directed their action. Denner (1970) found that normal readers were superior to problem readers of the same age in this type of logograph synthesis.

The link between operativity and reading comprehension has been investigated by Rawson (1969) and Stack and Murray (1976). In each study, stories were constructed that illustrated or reflected one or more of the operativity structures. In the Stack and Murray study these were class inclusion, complementary classes, multiplication of classes, seriation, and multiplication of relations—at the concrete operational level—and mechanical equilibrium, probability, proportionality, combinatorial reasoning, and correlation—at the formal operational level. Like the earlier example of the fat man and his pie, the chapters in the comprehension story each had one of these concrete or formal operativity structures embedded in it. Comprehension was measured (Rystrom, 1970) by measures of vocabulary, statement recall, statement sequence, inference, role taking, and so forth. The overall correlation between operativity, as measured independently in the traditional way, and comprehension was .69. The correlation was reduced to .51 when IQ and vocabulary (as measured by vocabulary subtest of the Gates-MacGinitie) were partialled out. There were as well significant low positive correlations between each operativity measure and its comprehension counterpart. Performance on the concrete operativity tasks exceeded comprehension of the passages in which these operativity structures were embedded while, surprisingly, comprehension of the

formal operational passages exceeded performance on the formal operational tasks themselves. Since the factor structures of the operativity task performance and the comprehension task performance were quite different and theoretically sensible, it does not seem that the achievement of this study was simply a measure of operativity in two modes, the traditional assessment paradigm and comprehension of a story passage.

The implication of this research is that there is a clear link between specific operativity features and specific aspects of reading comprehension, and that, at least at the concrete operational level, operativity may be a sufficient, if not necessary, condition for reading. What the exact contribution of operativity is to the reading process is not clear, but it is possible, and even likely, that it does not influence the process itself as much as it influences the child's understanding of the process. It may only allow the child to be more reflective about what he is doing when he reads, and to appreciate the inherent logic and structure in the activity of reading. Children who read without operativity to support the activity may be like children who can count accurately but do not have number conservation. Decoding certainly may occur in an automatic nonreflective manner and operativity presumably contributes very little to it; however, decoding with intelligence probably depends upon the Piagetian operational structures. Certainly, the language of reading instruction itself, emphasizing as it does the concept of rules, and classroom reading exercises of sorting, ordering graphemes and phonemes may be expected to depend upon operativity. Indeed, in studies of the correlation between operativity and reading, the measure of reading is often such classroom or standardized exercises. The concepts of rule, arbitrariness, and exceptions to rules, which underly phonic structure and the teaching of it, are concrete operational constructions. Piaget's work on moral development, justice, and the child's concept of the rules of a game indicate that these notions have a developmental history like that of operational intelligence. In sum, the theory should hold that in principle written material is a part of the world to be interpreted and understood by the same operations that Piaget has claimed support the child's understanding of the many things in the child's world that Piaget and his colleagues have already researched. Letter, word, and sentence are as valid concepts and appropriate objects for operational processing as are the concepts of number, weight, causality, time, amount, and velocity. These and other linguistic concepts are amenable to the operations of addition and multiplication of classes and relations that constitute the concrete operational structures as well as the operations of the INRC group of the formal operational period.

As the Stack and Murray study implied, the relationship between language (reading) and operativity may be somewhat

different at the formal operational level from what it was at the concrete level. Formal operations are hypothesized by Piaget to describe and explain some additional qualitative changes in the child's reasoning that occur around the ages of eleven and twelve years. At this time children seem for the first time to reason hypothetically, that is to imagine a situation different from the one that actually exists and to reason from it; they also show initial competence in concepts of probability, and proportion, and give evidence of combinatorial reasoning in which they seem able to hold all things constant but one in their problem solving. The notion here is that language (and by implication, reading) may be a necessary, but insufficient, condition for the development of formal operations (e.g., Elkind, Barocas, and Rosenthal, 1968). There is evidence that the written compositions and reading comprehension of formal operational children, or at least of children of that age, show clear and scoreable signs of formal operational thought (Lovell, 1971; Hallam, 1967; Goldman, 1964; Stack and Murray, 1976). This may indicate that the mental manipulation provided in language and reading is critical to the development of formal operational thought since it provides the medium for hypothetical and combinatorial reasoning. Of course, Piaget insists that formal operations, although not quite as general or universal as was once thought, have their developmental roots not in language but in the fusion of reversibility by negation and reversibility by reciprocity into one system (the INRC group).

Specific recommendations from Piaget's theory for reading instruction should wait for more explicit determinations of the relationships between operativity and decoding and comprehension. Still the evidence that exists suggests that the recommendations that might eventually come from the theory would include a general recommendation to delay reading instruction until the advent of concrete operational thought because these operations probably contribute to a more reflective appreciation of the reading task and because instruction in reading otherwise, while possible, is difficult and takes time from instruction that would contribute to the development of operativity itself. However, rather than wait for the match to occur between cognitive structure and the structure of the decoding task, schools could do some things to facilitate the match and bring about cognitive readiness for reading.

It would be reasonable to train children on the cognitive and perceptual operations that are related to reading. Rather than fit the child's mind to the task, however, it is also reasonable to fit the task to the child. Modifications in the reading task consistent with the theory already exist. The initial teaching alphabet (ita), for example, eliminates the need for a double classification competence of concrete operations since each phoneme has a unique grapheme. The language experience approach to reading instruction consti-

tutes a significant modification of the reading task since the child reads his own language and since the task itself is based upon the content and structure of the child's experiences. The match between cognitive structure and the task exists by definition in this instructional innovation.

In sum, the principal recommendation from Piaget's theory for reading instruction hinges upon a conceptualization of reading readiness in terms of operativity. A subsequent modification of the child's cognitive structure by an instructional technique, or the modification reading task itself through a curricular reform to facilitate a match between the structures of each, is a recommendation that is also consistent with the theory.

References

Beilin, H. "The Training and Acquisition of Logical Operations," in M.F. Rosskopf, L.P. Steffe, and S. Tabock (Eds.), *Piagetian Cognitive-Development Research and Mathematical Education*. Washington, D.C.: National Council of Teachers of Mathematics, 1971.

Beilin, H. "Inducing Conservation through Training," in G. Steiner (Ed.), *Psychology of the 20th Century*, Volume 7, *Piaget and Beyond*. Bern: Kinder.

Brainerd, C.J., and T.W. Allen. "Experimental Inductions of the Conservation of 'First-Order' Quantitative Invariants," *Psychological Bulletin*, 75 (1971), 128-144.

Brekke, B.W., J.D. Williams, and S.D. Harlow. "Conservation and Reading Readiness," *Journal of Genetic Psychology*, 123 (1973), 133-138.

Briggs, C., and D. Elkind. "Cognitive Development in Early Readers," *Developmental Psychology*, 9 (1973), 279-280.

Carroll, J.B. "The Nature of the Reading Process," in H. Singer and R.B. Ruddell (Eds.), *Theoretical Models and Processes of Reading*, Second Edition. Newark, Delaware: International Reading Association, 1976.

Denner, B. "Representational and Syntactic Competence of Problem Readers," *Child Development*, 41 (1970), 881-887.

DeVries, R. "Relationships among Piagetian, IQ, and Achievement Assessments," *Child Development*, 45 (1974), 746-756.

Elkind, D., R. Barocas, and B. Rosenthal. "Combinatorial Thinking in Adolescents from Graded and Ungraded Classrooms," *Perceptual and Motor Skills*, 27 (1968), 1015-1018.

Elkind, D., and J.A. Deblinger. "Perceptual Training and Reading Achievement in Disadvantaged Children," *Child Development*, 40 (1969), 11-19.

Elkind, D., J. Horn, and G. Schneider. "Modified Word Recognition, Reading Achievement, and Perceptual Decentration," *Journal of Genetic Psychology*, 107 (1965), 235-251.

Elkind, D., M. Larson, and W. Van Dorninck. "Perceptual Decentration Learning and Performance in Slow and Average Readers," *Journal of Educational Psychology*, 56 (1965), 50-56.

Farnham-Diggory, S., and M. Bermon. "Verbal Compensation, Cognitive Synthesis, and Conservation," *Merrill-Palmer Quarterly*, 1968, 215-228.

Furth, Hans. *Piaget for Teachers*. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Gibson, E.J. "Perceptual Learning in Educational Situations," in R.M. Gagne and W.J. Gephardt (Eds.), *Learning Research and School Subjects*. Itasca, Illinois: F.E. Peacock, 1968.

Goldman, Ronald. *Religious Thinking from Childhood to Adolescence*. New York: Seabury Press, 1968.

Goldschmid, M., and P. Bentler. "Concept Assessment Kit Conservation Manual." San Diego: Educational and Industrial Testing Service, 1968.
Goodnow, J.J., and G. Bethon. "Piaget's Tasks: The Effects of Schooling and Intelligence," *Child Development*, 37 (1966), 573-582.
Grant, Janet. "Piagetian Development and Learning to Read," unpublished master's thesis, University of London, 1972.
Hallam, R.N. "Logical Thinking in History," *Educational Review*, 19 (1967), 183-202.
Heatherly, A.L. "Attainment of Piagetian Conservation Tasks in Relation to the Ability to Form Hypotheses As to the Probable Content of Story Material among First and Second Grade Children," doctoral dissertation, University of Virginia, 1972.
Kaufman, A.S., and N.L. Kaufman. "Tests Built from Piaget's and Gesell's Tasks as Predictors of First Grade Achievement," *Child Development*, 43 (1972), 521-535.
Lovell, K. "A Follow-Up Study of Inhelder and Piaget's 'The Growth of Logical Thinking,'" *British Journal of Psychology*, 52 (1961), 143-154.
Lovell, K., D. Shapton, and N.S. Warren. "A Study of Some Cognitive and Other Disabilities in Backward Readers of Average Intelligence As Assessed by a Nonverbal Test," *British Journal of Educational Psychology*, 34 (1964), 58-64.
Mermelstein, E., and L.S. Shulman. "Lack of Formal Schooling and the Acquisition of Conservation," *Child Development*, 38 (1967), 39-52.
Rausher, S.R. "The Relationship between Achievement on Piagetian Conservation and Spatial Measures and Reading Readiness," doctoral dissertation, New York University, 1970.
Rawson, H.I. "A Study of the Relationships and Development of Reading and Cognition," unpublished doctoral dissertation, University of Alberta, 1969.
Rystrom, R. "Toward Defining Comprehension: A First Report," *Journal of Reading Behavior*, 2 (1970), 56-73.
Scott, R. "Social Class, Race, Seriation, and Reading Readiness: A Study of the Relationship at Kindergarten Level," *Journal of Genetic Psychology*, 115 (1969), 87-96.
Smith, F. *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read*. New York: Holt, Rinehart and Winston, 1971.
Stack, W., and F.B. Murray. "Operativity and Reading Comprehension," paper presented at the Sixth Annual Symposium of the Jean Piaget Society, June 1976.
Waller, T.G. *Think First, Read Later!* IRA Series on the Development of the Reading Process. Newark, Delaware: International Reading Association, 1977.

Elementary school teachers ask their students many questions. Do those questions help children develop their thinking? What kinds of questions should be asked? It appears that most classroom inquiries concern facts. Higher levels of questioning also are possible. Robert Ruddell presents a framework for instruction that incorporates comprehension skills at three levels. This framework, in combination with concern for specific questioning strategies and awareness of who is asking or responding, will enable teachers to analyze and improve their questioning in order to help children develop comprehension abilities.

Developing comprehension abilities: implications from research for an instructional framework

ROBERT B. RUDDELL
University of California at Berkeley

A major objective of the instructional program is the development of reading and listening comprehension abilities that will enable the child to effectively derive, interpret, and apply meaning. These abilities directly influence the success achieved by the individual throughout the primary, intermediate, junior high, and high school levels and become even more critical as a student enters college or a trade, becomes employed, and assumes personal life responsibilities. Clearly, comprehension abilities are of vital importance to personal success.

In the following discussion we will first briefly explore logical analysis and empirical research related to the nature of the comprehension process. Second, an instructional framework will be formulated to account for comprehension levels and competencies, giving attention to questioning strategies of value to the practitioner in the classroom.

Both logical analysis and empirical study indicate that the comprehension process involves a cluster of related mental abilities rather than a single mental skill (Ruddell, 1974). The work of Bloom et al. (1956) is one of the best known examples of logical analysis in the cognitive area. This work views cognition from the standpoint of five levels: remembering, reasoning, problem solving, concept formation, and creative thinking. Although the more abstract and advanced levels appear to include requisite mental processes from the simpler level, there is obvious overlap in the mental processes.

Barrett (1968)—utilizing the research of Bloom (1956), Sanders (1966), Letton (1958), and Guzak (1965)—designed a

taxonomy of cognitive and affective dimensions that accounts for various levels: literal comprehension, reorganization, inferential comprehension, evaluation, and appreciation. Within each category, Barrett identifies specific skill objectives to be used for instructional purposes. Though based primarily on a logical analysis of the comprehension process, such a taxonomy can be of value to the classroom teacher in formulating questions and in initially assessing the relationship between cognitive levels and difficulty of questions.

Numerous empirical studies provide information on comprehension levels and skill competencies. In summarizing approximately thirty studies, Lennon (1960) identifies several components of reading comprehension:

1. A general verbal factor closely related to vocabulary knowledge.
2. Comprehension of explicitly-stated materials, the understanding of literal meanings, and the ability to follow directions.
3. The comprehension of implicit or latent meanings such as the ability to draw inferences, to predict outcomes, and to perceive a hierarchical arrangement of ideas within a selection labeled "reasonsing in reading."
4. Appreciation of the work by "seeing the intent or purpose of an author, judging the mood or tone of a selection [and] perceiving the literary devices by means of which the author accomplishes his purposes."

Davis' review of comprehension research (1971) reveals a range of comprehension competencies, including knowledge of and recalling word meanings; the ability to grasp detailed facts; drawing inferences about the meaning of a word from context; following the structure of a passage; weaving together ideas in the content; drawing inferences from the content; identifying a writer's techniques; recognizing a writer's purpose, attitude, tone, and mood; and ability to consider a passage objectively without being overwhelmed by personal experiences and feelings. Moreover, Davis found that the knowledge of word meaning and the ability to draw inferences from the content are two factors most critical to the comprehension process.

Development of the comprehension process relies heavily on what Taba has termed "cognitive commerce," that active cognitive process occurring when the child interacts with the environment (Taba, 1965; Taba and Elzey, 1964). The question is a basic and commonly accepted tool used to stimulate thinking as well as to enhance the cognitive process and comprehension ability. Questions range from a focusing type—devised to establish a mental set or purpose for reading—to a raising type—designed to obtain additional information at a higher level of comprehension. Questions serve to clarify facts or ideas and to encourage the child to

involve higher thinking skills—to consider the "why" of a process or phenomenon. As Freedle and Carroll (1972) point out, questions may serve to control "the reader's expectancies concerning what may be critical in the text, and thus of his attention to particular classes of details." In brief, questions allow a "guided exploration" approach to instruction by stimulating children to search for specific information clues, establish cause and effect relationships, and derive inferences. Skillful questioning strategies not only provide direct feedback to the child by indicating the appropriateness of a response but also provide direction for continued hypothesis building during problem solving.

The research by Taba (1964, 1965); by Wolf, Huck, and King (1967); and by Almy (1966) supports the view that questions and questioning strategies can be valuable instructional tools useful in developing the child's cognitive abilities. Unfortunately, it appears that teachers have had little opportunity to develop either competence or confidence in understanding or using questioning strategies or to seriously consider the importance of comprehension levels and skill competencies. Research by the present writer revealed that in the primary grades approximately 70 percent of teacher questions are at the factual or literal comprehension levels (Ruddell and Williams, 1972). Guzak (1967) substantiates these findings with students in the second, fourth, and sixth grades where his investigations revealed that 70 percent of the questions are concerned, for the most part, with "factual make-up of stories." Guzak believes that such questions actually "lead the student away from basic literal understanding of story plots, events, and sequences." Therefore, when devising factually oriented questions, which may really be trivial-fact questions, teachers must constantly be aware of the danger that the child may miss the understanding of the broader text.

The previous logical analysis and empirical research on the nature of the comprehension process was used to assist the formulation of the following comprehension framework (Ruddell, 1974). This framework accounts for the interaction of three comprehension levels identified as factual, interpretive, and applicative, and specific skill competencies ranging from details through problem solving.

The comprehension levels vary along a continuum from recall of information to utilization of information in new situations. As viewed within this framework, the factual level relates directly to the factual content—information and ideas related to the reader or listener. To comprehend material at this level requires identification and recall of information. An important distinction between this level and the interpretive and applicative levels is that, at the factual level, the individual is not required to cognitively manipulate or control the details, sequence of ideas, and so on. The response

Skill Competencies	*Comprehension Levels*		
	Factual	*Interpretive*	*Applicative*
1. Details			
a. Identifying	✓	✓	
b. Comparing	✓	✓	✓
c. Classifying		✓	✓
2. Sequence	✓	✓	✓
3. Cause and Effect	✓	✓	✓
4. Main Idea	✓	✓	✓
5. Predicting Outcome		✓	✓
6. Valuing			
a. Personal judgment	✓	✓	✓
b. Character trait identification	✓	✓	✓
c. Author's motive identification		✓	✓
7. Problem Solving			✓

may include valuing at a very basic level but would require no substantiation or support for the yes or no answers.

At the interpretive level, however, the reader or listener must modify and manipulate the content by analyzing, reconstructing, and inferring relationships using skill competencies such as details, sequence, cause and effect, main idea, and prediction of outcome. At this level, the individual must cognitively manipulate the ideas and information as reflected in his or her response although substantiation or support for the inferences derived may not be extensive. Valuing at the interpretive level may reflect inference based on personal experiences and judgment, character trait identification, and identification of the author's motive.

Finally, the applicative level of comprehension requires the reader or listener to transform, utilize, and apply the information obtained by providing—in a new situation—evidence of application through either verbal substantiation or direct manipulation of elements in the environment.

Previous instructional experience has familiarized most teachers with the skill competencies identified in the instructional framework. The unique function of the framework as presented here, however, lies in incorporating each skill competency at a particular comprehension level. It is important that teachers develop an awareness of the range of competencies across levels as questions are formulated. To be useful to the classroom teacher, the foregoing research discussion and instructional conceptualization must be applied to the reality of the instructional setting. A brief excerpt from *Charlotte's Web*, with accompanying questions, will illustrate identification of comprehension levels and skill competencies in this regard.

The reader will recall that at the point of the following excerpt, Wilbur has become even more rotund, and Charlotte is concerned about his future in the probable form of bacon in the Zukerman's smokehouse:

> On foggy mornings, Charlotte's web was truly a thing of beauty. This morning each thin strand was decorated with dozens of tiny beads of water. The web glistened in the light and made a pattern of loveliness and mystery, like a delicate veil. Even Lurvy, who wasn't particularly interested in beauty, noted the web when he came with the pig's breakfast. He noted how clearly it showed up and he noted how big and carefully built it was. And then he took another look and he saw something that made him set his pail down. There, in the corner of the web, neatly woven in block letters, was a message. It said:
>
> SOME PIG!
>
> Lurvy felt weak. He brushed his hand across his eyes and stared harder at Charlotte's web (White, 1952).

The following questions and answers illustrate the use of comprehension levels and skill competencies.

1. What was written in Charlotte's web?
   Answer: SOME PIG!
   Comprehension Level/Skill Competency:
   Factual—Identifying details.
2. Why did Lurvy feel weak?
   Answer: Interpretation suggests that Lurvy believed it a rather unusual phenomenon to find the words SOME PIG! woven into a spider's web, and his actions express astonishment and surprise.
   Comprehension Level/Skill Competency:
   Interpretive—Cause and effect.
3. If you had been the first person to see Charlotte's web that morning, how would you have reacted?
   Answer: Responses will vary.
   Comprehension Level/Skill Competency:
   Applicative—Predicting outcomes.

A note of caution in regard to utilizing comprehension levels and skill competencies must be observed, especially when designing questions at the interpretive and applicative levels. Formulating questions for instructional purposes requires that the teacher actually examine the story to determine whether or not a specific question is answered directly in the text. For example, in relation to the foregoing excerpt, consider the following question: "Why was Lurvy in the barnyard that morning?" Based on the "why" nature of the question, the teacher might assume immediately that an appropriate response would necessitate the use of the interpretive level; however, close examination of the text reveals the sentence: "Even Lurvy, who wasn't particularly interested in beauty, noted

the web when he came with the pig's breakfast." In other words, if the answer to the question is stated in the text and therefore requires only informational recall rather than manipulation of information to arrive at possible interpretation or application, the question is at the factual level and, in this case, uses a cause and effect competency. By developing a greater sensitivity to the significance of comprehension levels and skill competencies, teachers can effectively employ questions at the interpretive and applicative levels to enhance the classroom comprehension program through reading, listening, and discussion activities.

As discussed earlier, verbal interaction is frequently employed in the classroom setting but consists primarily of factual or literal comprehension questions and responses among teachers and children. By improving awareness of various questioning strategies teachers can greatly increase the potential for developing comprehension levels and skill competencies when involving students in verbal interaction. Seven questioning strategies are described briefly in the following discussion (Ruddell, 1974; Taba, 1965; Taba and Elzey, 1964):

1. *Focusing* is a strategy used to initiate discussion or to refocus on an initial question. In a sense, a focusing question or response helps develop a mental set for verbal interaction. For example, during discussion of Hoban's *Charley the Tramp*, the teacher might ask: "Well, Charley certainly seemed to be a happy beaver; why do you think he left home?" This constitutes a focusing question at the interpretive level, involving a cause and effect competency.

2. *Ignoring* is a strategy everyone experiences periodically. Ignoring essentially occurs when a question or response is not accepted, and this rejection may result from such factors as time pressure or negative attitude toward an individual.

3. *Controlling* is a strategy whereby a single individual dominates the discussion, providing little opportunity for others to interact verbally. For example, the teacher might control the discussion by posing questions such as the following at the interpretive level, involving personal judgment:

That was a really good story, wasn't it?

I think Willie made the right choice, don't you?

Who would you like to be in this story? Would you like to be Tony?

4. *Receiving* is a strategy by which a question or response is acknowledged but without any elaboration whatsoever. It is interesting to note that receiving can be either positive or negative—depending upon the actual response, the vocal tone, or the facial expression. Repeating the child's statement (as classroom teachers so frequently do) provides the most customary illustration of the receiving strategy. Or the teacher might say: "Right! Good thinking.

That's good thinking." Such a response constitutes receiving at the interpretive level, involving personal judgment.

5. *Extending* is a strategy used to elicit additional information on the same subject and at the same comprehension level. The following teacher questions or comments are examples of extending:

Is there anything else you could say about it now?
Keep going, that sounds interesting.
Who has another idea about that?

6. *Clarifying* is a strategy in which the question or response explains or redefines previous information. Clarifying differs from extending because this strategy encourages an individual to deal with a prior and unclear statement. Clarifying may also utilize questions designed to obtain information on the same subject at a lower comprehension level as well as on the same subject at the same comprehension level. For example, the following teacher comments represent clarifying:

Oh, I see what you mean, it could have been . . . so and so.
Could you explain your point again, please.
Do you mean that . . . so and so?

7. *Raising* is a strategy characterized by an attempt to solicit additional information on the same subject but at a higher comprehension level. In other words, raising constitutes a shift upward from the factual to the interpretive level or from the interpretive to the applicative level. Examples of this strategy would include questions such as the following:

Why would you especially like to be Alexander?
If that did happen, what kind of a problem do you think you'd have?
Why do you like this ending best?

Of the seven questioning strategies identified above, four have proved to be most significant as vehicles for effectively developing the child's comprehension abilities. Focusing enables the teacher to immediately establish a mental set, a purpose for reading. Extending allows the teacher to elicit additional information on the same subject at the same comprehension level. Clarifying enables the teacher to encourage returning to a previous response for further clarification, explanation, or redefinition. Raising allows the teacher to obtain additional information on the same subject but at a higher comprehension level.

One approach to the improvement in awareness and sensitivity to comprehension levels and skill competencies is the utilization of the previously discussed levels of comprehension with the specific questioning strategies which serve as the basis of viewing verbal interaction in the classroom. Many systems of interaction analysis are extremely time-consuming and incorporate extensive definitions and categories with the result of minimal

teacher interest. In the attempt to alleviate this obvious difficulty, the present writer has incorporated the three comprehension levels and seven questioning strategies into an observational instrument for the classroom. The levels and strategies are directly related to the speaker and the function of the interaction. Examination of the following chart indicates that the teacher needs to consider four specific dimensions when employing the simplified interaction analysis.

1. Who talks—the teacher or the child?
2. The function of the interaction—the question or the response.
3. The comprehension level—factual, interpretive, or applicative.
4. The questioning strategy—focusing, ignoring, controlling, receiving, extending, clarifying, or raising.

1	*2*	*3*	*4*
Who Talks	*Function*	*Levels of Comprehension*	*Strategies*
Teacher	Question	Factual	Focusing
Child	Response	Interpretive	Ignoring
		Applicative	Controlling
			Receiving
			Extending
			Clarifying
			Raising

Using the following self-analysis observational scale (Ruddell, 1974), the classroom teacher can initiate self-analysis using a tape recording of a classroom reading lesson or storytelling activity.

Such self-reflection will indicate not only which comprehension levels and questioning strategies are employed but also the degree of effectiveness or ineffectiveness these interaction techniques permit during an instructional discussion based on the fit between teacher question and child response. A central objective in conducting such analysis is the development of instructional sensitivity to the use of questioning levels and strategies in daily instruction regardless of content area.

Verbal interaction accounting for factual and interpretive levels and questioning strategies between teachers and children in the classroom setting is illustrated in Tables 1 and 2. These data are derived from videotapes recorded during a research investigation of a literacy teaching model directed by Ruddell (1972). The questioning levels and strategies of 24 primary grade teachers and the response levels and strategies of 144 primary grade children were studied.

Self-analysis Record of Classroom Interaction

Date:
Teacher:
Lesson:
Description of Teaching Situation:

Focusing Question:

1	2	3	4
F Foc Ign T Q I Cont C R Rec A Ext Clar Rais	F Foc Ign T Q I Cont C R Rec A Ext Clar Rais	F Foc Ign T Q I Cont C R Rec A Ext Clar Rais	F Foc Ign T Q I Cont C R Rec A Ext Clar Rais
5	**6**	**7**	**8**
F Foc Ign T Q I Cont C R Rec A Ext Clar Rais	F Foc Ign T Q I Cont C R Rec A Ext Clar Rais	F Foc Ign T Q I Cont C R Rec A Ext Clar Rais	F Foc Ign T Q I Cont C R Rec A Ext Clar Rais

A brief survey of the table reveals that teacher questions at the factual level account for 68.2 percent of the total questions, whereas teacher questions at the interpretive level account for only 31.8 percent of the total. Teachers, thus, pose approximately twice as many factual as interpretive questions. The extending strategy is most frequently employed (57.5 percent) when asking questions at the factual level and the raising strategy (49.7 percent) when asking questions at the interpretive level. Further, Table 1 indicates that although the raising strategy occurs most often in conjunction with the interpretive level, the focusing strategy and the extending strategy combine to represent 70.9 percent of all strategies used.

Table 2 reveals that child responses at the factual level account for 85.9 percent of the total, whereas child responses at the interpretive level account for 14.1 percent of the total. Obviously, a number of the teacher's questions at the interpretive level (31.8 percent, Table 1) were not handled effectively by the child's responses at the interpretive level (14.1 percent, Table 2). In addition, child responses at the factual level (85.9 percent, Table 2) far outweigh teacher questions at the factual level (68.2 percent, Table 1).

Table 1
Teacher Questioning Levels and Strategies
Taxonomy Analysis—Video Recording

Questioning Level	Focusing Mean	Focusing %	Controlling Mean	Controlling %	Receiving Mean	Receiving %	Clarifying Mean	Clarifying %	Extending Mean	Extending %	Raising Mean	Raising %	Total	% By Level
1. Factual	29.06	26.00	4.93	4.40	8.06	7.10	5.67	5.00	65.00	57.50			112.72	68.20
2. Interpretive	12.87	24.40	3.44	6.50					10.10	19.20	26.19	49.70	56.60	31.80
Total	41.93		8.37		8.06		5.67		75.10		26.19		165.32	
% By Strategy		25.30		5.10		4.90		3.40		45.60		15.70		100.00

Table 2
Child Response Levels and Strategies
Taxonomy Analysis—Video Recording

Response Level	Focusing Mean	Focusing %	Controlling Mean	Controlling %	Receiving Mean	Receiving %	Clarifying Mean	Clarifying %	Extending Mean	Extending %	Raising Mean	Raising %	Total	% By Level
1. Factual	.09	.05			57.03	33.10	.09	.05	115.40	66.80			172.61	85.90
2. Interpretive									25.75	91.90	2.27	8.10	28.02	14.10
Total	.09				57.03		.09		141.15		2.27		200.63	
% By Strategy		.05				28.40		.05		70.40		1.10		100.00

The apparent discrepancy between teacher questions and child responses might be explained in the observation that 15.7 percent of teacher questions utilized the raising strategy whereas only 1.1 percent of the child responses reflected the same strategy. Therefore, it appears that the children—being unable to handle teacher questions utilizing a raising strategy—responded, instead, by using an extending strategy; that is, the extending strategy occurs at a frequency of 70.4 percent in child responses in contrast to only 45.6 percent in teacher questions. In conjunction with this explanation, Table 2 reveals that child responses at the interpretive level using the extending strategy constitute 91.9 percent whereas teacher questions at the interpretive level using the extending strategy constitute only 19.2 percent. The child responses at the factual level using the extending strategy constitute 66.8 percent, whereas teacher questions at the factual level using the extending strategy constitute only 57.5 percent.

The findings presented in these two tables illustrating verbal interaction in the classroom suggest that comprehension levels and questioning strategies used by the teacher must be designed with sensitivity to child response levels and strategies. It is extremely important that instruction not depend so heavily on factual questions that merely require the child to recall or identify story details. Instead, teachers must strive to develop a guided exploration approach by emphasizing questions at the interpretive and applicative levels using focusing, extending, clarifying, and raising strategies appropriate for the child's level of cognitive development. Combining these four strategies with a thorough understanding of comprehension levels and skill competencies will provide teachers with an invaluable instructional tool—a tool that can greatly assist in designing a comprehension program to match and enhance the child's cognitive development. The effort a teacher expends in building the comprehension program will be reflected directly in students' abilities to effectively derive, interpret, and apply meaning from oral and written communication experiences encountered throughout life.

References

Almy, Millie. *Young Children's Thinking*. New York: Teacher's College Press, Columbia University, 1966.

Barrett, Thomas C. "Taxonomy of Cognitive and Affective Dimensions of Reading Comprehension," in Theodore Clymer, "What Is 'Reading?' Some Current Concepts," *Innovation and Change in Reading Instruction*, Sixth-Seventh Yearbook, National Society for the Study of Education, Part II, Helen M. Robinson, Ed. Chicago: University of Chicago Press, 1968, 19-23.

Bloom, Benjamin S., et al. (Eds.). *Taxonomy of Educational Objectives: Handbook 1, Cognitive Domain*. New York: David McKay, 1956.

Davis, Frederick B. "Psychometric Research on Comprehension in Reading," in F.B. Davis (Ed.), *Targeted Research and Development Program in Reading with Emphasis on Models*. Washington, D.C.: Project No. 2, U.S. Department of Health, Education and Welfare, Contract No. OEC-0-70-4790 (508), Project No. 0-9030 (1971), 8-3: 8-60.

Freedle, Roy O., and John B. Carroll. "Language Comprehension and the Acquisition of Knowledge: Reflections," *Language Comprehension and Acquisition of Knowledge*. Washington, D.C.: V.H.Winston and Sons, 1972, 359-368.

Gusak, Frank J. "Teacher Questioning and Reading," *Reading Teacher*, 21 (1967), 227-234.

Gusak, James. "Reading Comprehension Solicitation Response Inventory," unpublished manuscript, University of Wisconsin, 1965.

Lennon, Roger T. "What Can Be Measured?" in Russell Stauffer (Ed.), *The Role of Tests in Reading*, Proceedings of Annual Education Conference, 9. Newark: University of Delaware Press, 1960, 67-80.

Letton, Mildred C. "Evaluating the Effectiveness of Teaching Reading," in Helen M. Robinson (Ed.), *Evaluation of Reading*, Supplementary Education Monographs, No. 88. Chicago: University of Chicago Press, 1958, 76-82.

Ruddell, Robert B. *Reading-Language Instruction: Innovative Practices*. Englewood Cliffs, New Jersey: Prentice-Hall, 1974, 361-408.

Ruddell, Robert B., and Arthur C. Williams. *A Research Investigation of a Literacy Teaching Model: Project* DELTA. Washington, D.C.: U.S. Department of Health, Education and Welfare, Office of Education, EDPA Project No. 005262 (1972).

Russell, David H. "Comprehension: Literal and Interpretive," in Robert B. Ruddell (Ed.), *The Dynamics of Reading*. Boston: Ginn, 1970, 160.

Sanders, Norris M. *Classroom Questions*. New York: Harper and Row, 1966.

Taba, Hilda. "The Teaching of Thinking," *Elementary English*, 42 (1965), 534-542.

Taba, Hilda, and Freeman F. Elzey. "Teaching Strategies and Thought Processes, *Teachers College Record*, 65 (1964), 524-534.

White, E. B. *Charlotte's Web*. New York: Harper and Row, 1952, 77.

Wolf, W., C. S. Huck, and M. L. King. *Critical Reading Ability of Elementary School Children*. Washington, D.C.: U.S. Department of Health, Education and Welfare, Project No. 5, 1040 (1967).

What is the content of children's reading books? Is the content racist and sexist? Do the stories reflect the realities of community, family, and child life; do they motivate children to read? Do children in the United States read stories that differ from those children read in Europe?

Gaston Blom has investigated these and other questions about children's reading textbooks because he feels that *what* children are taught to read matters more than *how* they are taught to read. He thinks that reading that serves children by providing useful knowledge, teaching social values, allowing emotional expression, and motivating further reading will also improve the children's reading skills. In this selection, Dr. Blom summarizes studies of content of children's books in the United States and other countries.

The role of content in the teaching of reading

GASTON E. BLOM
Michigan State University

It has been recognized for a long time that motivation is a powerful influence on learning to read and on the further development and practice of reading. While this is accepted intuitively, the how and process of reading, rather than the why and purpose, usually receive the greatest attention of teachers and reading specialists. Yet most people would agree that reading without meaning and comprehension is not reading. However, there is considerable evidence that schools are more concerned about achievement test literacy than they are about practiced literacy. Achievement test literacy means adjustment to school culture and demands at the expense of the many values reading can provide. To some extent, schools are responding to the pressures of parents and of society to develop reading skills (in children and adults) which often ignore the many adaptive and pleasurable functions of reading. As a consequence, emphasis on learning to read has been greater than reading to learn. The average child between the ages of six and eleven spends no time, or very little, at voluntary reading outside of school (Roberts and Baird, 1971). Evidence suggests that the incorporation of functions of reading into instructional reading programs still receives insufficient attention in this country.

Motivation to read is, of course, a complex issue. Some of its elements are presented in past and current books on the reading process. Gibson and Levin (1975) refer to intrinsic motivation in children to get information about things, people, places, and events; to the importance of encouraging the free choice of reading materials; and to the significance of role models (teachers and parents) and of environments where written materials are available

and used. These authors review their impressions and some of the studies on the content of instructional materials which point to alienness, redundancy, awkward linguistic style, and lack of emotional impact.

This paper will focus on studies of the content of reading materials and their influence on reading and other behaviors of children. The many other motivational influences on the reading process, as well as cognitive, linguistic, instructional, and biological factors, will not be examined. However, the author acknowledges that reading is a complex behavior which is one of the functions of language and that all language functions are influenced by a large number of intrinsic and extrinsic factors. So while content is important and is important in relation to the many factors influencing reading, it should be viewed in the context of a complex behavioral activity. The bias of this author is that reading and other language behaviors are best viewed as a developmental naturalistic process.

Content of American primers

A review of articles which appeared in the popular and educational press up to 1968 (Blom et al., 1968) called attention to many inappropriate elements in the content and format of first grade reading textbooks currently in use in the United States. These included:

1. lack of recognition of ethnic, national, and cultural differences;
2. relative absence of realistic information about the world and of knowledge in general;
3. restriction to middle-class and suburban life;
4. use of limited vocabulary and unnatural language;
5. contradictions to the everyday experiences of children by avoiding difficulties and problems;
6. dearth of moral and ethical messages;
7. lack of awareness of sex differences in interests;
8. repetition of Pollyannish family activities of low interest appeal.

During the 1970s, major attention has focused on the content of primers which reflect sexism (Key, 1971; Frasher and Walker, 1972; Graebner, 1972; National Organization for Women, 1972; Britton, 1973; Tittle, 1973; Gaugh, 1974) and racism (Costo, 1970; Kane, 1970; Allen, 1971; Waite, 1972). The concern of these studies was the inappropriate, biased, and potentially harmful effects of content in socializing children to our culture. In addition, the alienness of content to many groups of children who did not come from intact, white, middle-class, suburban, Anglo-Saxon families might have a negative motivational influence on reading.

The studies on sex stereotyping pointed to fewer female characters in stories; boys as intelligent, competent, achieving, competitive, and active; and girls as passive, docile, incompetent, and spectators. Numerous occupational roles were displayed for men and very limited and restricted ones for women. Tittle (1973) specifically examined eight achievement test batteries for sex bias and found a preponderance of male nouns and pronouns as well as sex stereotypical roles for men and women. Graebner (1972) indicates that recent editions of basal readers have been modified slightly with societal changes but still do not approach reality.

While there has been some improvement in the presence and portrayal of nonwhite Americans in children's textbooks (Allen, 1971; Waite, 1972), the paucity of Latin Americans and American Indians remains in the content of stories. Costo (1970) points out how resource books about American Indians contain inaccuracies and distortions, and do not recognize Indian contributions to our society nor their having been in America long before white settlers. Kane (1970) indicates that a white, protestant, Anglo-Saxon viewpoint is still portrayed in social studies textbooks.

Prior to 1968, very little attention had been given to systematic studies of the content of primers although much had been written about them. A few studies on the content of third grade reading textbooks (Child et al., 1946; Holmes, 1960; McClellan, 1961) pointed to their Pollyannish qualities and middle-class American values and attitudes. Since 1968, a number of studies by the same research group (Blom et al., 1970; Zimet, 1972) found similar attributes in first grade American reading textbooks that were analyzed using 103 motivational variables along 8 content dimensions (characters, character composition, predominant theme, age of activity, sex of activity, outcome of activity, environmental setting, and cultural composition).

The findings from the Blom and Zimet studies were obtained on a representative national sample of commonly used primers from twelve textbook publisher series. While some changes in content were noted in primers published after 1962, content still remained inappropriate in relation to sex roles, cultural differences, developmental interests of children, and depiction of real life. A statistical description of 1,307 stories provided a picture of what most American children read in first grade. A large proportion of stories (47 percent) had stereotyped themes of real life, with positive emotions, active play, and pets, often described as Pollyannish. Very few stories dealt with school (3 percent), pranks and humor (3 percent), or lessons from life (1 percent). There was considerable restriction in the range of character combinations, with children and animals (pets) comprising 23 percent and adults only 2 percent. Of the 1,307 stories, 1,161 included children, with the most typical distribution being a boy (77 percent) and a girl (72 percent), each

about six years of age, a younger sister (34 percent). Children older than six were rarely present (5 percent). Forty-five of the stories were rated as boy-girl activities with lack of clarity as to sex role and, statistically, boy activity stories ended in failure more often than did girl activity stories. While outcome usually was successful (64 percent), failure usually was not depicted in a realistic manner (i.e., what a child of a particular age could not do). Further analyses indicated that only 2 percent of the stories took place in an urban setting where most families live.

When so called multiethnic urban primer textbooks developed between 1965 and 1966, 820 stories in seven publication series were analyzed according to the previously described content dimensions. Results indicated that, in most ways, these new textbooks approximated the traditionally used primers described above. There was one exception. In outcome of activity in two of the seven multiethnic urban series, failure occurred more often than success and, in six of the seven series, help was much higher (13-26 percent) as compared to the national sample (9 percent). Each of the seven series dealt in its own way with the issue of ethnic composition of characters. In two series, 79 percent and 87 percent of the stories had white characters only and, in two additional series, whites only comprised the largest group of stories (49 percent and 40 percent). In three series, whites only were present in a small percentage of stories (0 percent, 4 percent, and 6 percent).

In a clinical analysis of some character types in one of these seven multiethnic urban reading series developed between 1965-1966, it was found that the black boy was portrayed as athletic, less intelligent, impulsive, distractible, and the object of humor. In contrast, the white boy was displayed as reflective, more intelligent, and socially secure. The fathers of these two boys also showed contrasting characteristics with the white father having greater economic resources, displaying more consistent "masculine" behaviors, and offering assistance to his black counterpart.

The interpretation of these data obtained through systematic analysis points to a gestalt. The activities are neutral and redundant without much content significance and variation. They are happy, family centered, and tend to be ambiguous or stereotyped as to sex role. A child is almost always with other children and is seldom alone. Older age children rarely appear as siblings and peers. In contrast, there tends to be a regressive pull through emphasis on family attachment and younger siblings, animal stories, anthropomorphized figures, and ambiguity in sex roles. The setting is most typically in the suburbs, rarely in the city, and in and around a home. Pets are amusing, cute, and frustrating nuisances.

This gestalt represents a striking divergence from the realities of community, family, and child life — from what is known about child development. Most children live in cities and are in

contact with other children of different skin color and national background. Family life is not so exclusively child centered, nor is it constantly happy and smiling. Parents punish and children display a range of affects. It is more typical to find older siblings and peers in greater frequency than the first grade primers demonstrate. Pets provide various meanings for children such as companions in adventure, exploration, and rough play and as methods for learning about biological happenings (pregnancy, birth, growth, infancy, weaning, toilet training, old age, death, illness, and injury). The stories obviously do not deal with these issues.

While publishers have responded to pressures for reading textbooks that include people of more than one ethnic background and that depict the urban setting, studies of seven multiethnic urban reading textbook series demonstrate that some of the series contain a few significant characters of ethnic background other than white Anglo-Saxon. When other ethnic groups are present, they frequently appear in traditional suburban oriented settings. Some series depicted a high incidence of failure as the outcome of a main activity. Only a few multiethnic series showed significant departures from the traditional content and format of first grade reading textbooks. Although the authors of multiethnic urban series have responded to an urgent social need to write textbooks depicting whites and nonwhites, for the most part they have not been successful in their attempts. The impression is that prejudicial attitudes and conflict between *what is* and *what should be* have been responsible for results that are not very satisfactory.

Responses of children to content

The crux of the studies on content is that reading textbooks used in the first grade have been inappropriate to the social, cultural, and developmental contexts of children and of the world in which they live. As such, they lack motivational interest. The assumption that motivational content influences reading acquisition and development is a reasonable one, but attempts to support this assumption scientifically have posed noteworthy methodological problems. Nevertheless, experiences of teachers and writers, as well as clinical evidence, support it.

Ashton-Warner (1963), for example, reported that improvement in the acquisition of learning to read English resulted when Maori children read their own written stories. This was striking when compared to results obtained from the use of traditional textbooks. These children wrote stories that depicted real life content themes with "organic" words of sex, hate, and fear. Some children had a secret vocabulary of special meaningful words which, when tapped, opened the door to reading.

The Fernald (1943) method of teaching reading, using a

kinesthetic approach, also has been adapted to having children write their own stories and learning to read them. Children make their own storybooks and have vocabulary boxes with "their own words." This is similar to learning experience approaches to reading, more often used as enrichment activities or in remediation and less used as a primary method of teaching reading.

Fader and McNeill (1968) have successfully used a saturation of the environment with reading approach to linguistically impoverished secondary school boys in urban and institutional settings. Paperbacks, magazines, and newspapers that were developmentally relevant and interesting were provided. These could be read in classrooms, checked out, borrowed, taken home, and lost. In addition, the authors got these children to write in their private journals a minimum of 1½ pages a day of any material they wished. Quantity and practice were considered crucial. These educational strategies combined desensitization, high motivational techniques, and practice.

Laubach and Laubach (1960) indicate from their experiences in teaching literacy in many languages that illiterate societies need to see a usefulness and purpose to the reading process. In essence, reading has to touch and influence their immediate lives. The Laubachs developed an English teaching program, based on word and picture association, for adults in the United States. It would seem apparent that similar motivation of usefulness exists for children whose environments may not communicate the significance and value of reading and for children who have difficulty grasping the importance of an immediate skill for long range goals. In fact, Merritt has developed a teaching of reading instruction based on the concept of functional purpose. Reading activities are designed to have practical purposes in the everyday lives of children, adolescents, and adults (Davies, 1973).

In addition, clinical impressionistic and empirical findings that content counts can be cited. However, to demonstrate content influence through objective systematic study presents numerable obstacles. Classroom field studies have these difficulties.

In 1962 a multiethnic urban first grade reading series was introduced into selected schools in a large midwest city (Whipple, 1963). The series was constructed to contain characters, themes, and language typical of multicultural neighborhoods. The teacher's manual accompanying the series emphasized social as well as skill objectives. Outcome measurements were obtained of word recognition, oral reading accuracy, and interest appeal and results were compared with those from matched schools using traditional all white books. The results clearly favored the multiethnic urban series (Whipple, 1964). Questions are raised about whether the initial favorable results were based upon the influence of enthusi-

asm, newness, and other attitudes. Follow-up studies indicating the maintained superiority of the multiethnic series have not been reported. Questioning the basis for these results should not be misinterpreted to mean that it is not highly important to present to children reading materials which contain a multicultural content. Informal discussions with nonmajority adults about their opinions of reading books having other than all white characters have nearly always been positive and enthusiastic.

Another field study compared the influence of a high boy interest (where boys and men are strong and intelligent and adventure and action predominate) basal reading series on reading achievement with a standard state adopted text considered low in boy interest (Stanchfield, 1973). The contrast text of low boy interest does not mean it was high in girl interest. First grade classes taught with this series and its method achieved significantly higher scores in reading for boys and girls than did control groups using a standard series. While the new series was high in boy interest, one should challenge the content which underplayed the role of girls and women and one could also question whether the results were due to factors other than content as mentioned before. In addition, there was evidence that special preparatory work with teachers may have biased the results (i.e., teacher factors).

The influence of experimental conditions on improved reading achievement has been reported by Zimet et al. (1974). Two first grade classes used new textbook series with contrasting content. One series consisted of revised standard primers, while content of the other series had impact and high interest. 1) The two populations of first graders made equal achievements in reading so that a specific role for textbook content could not be determined. No sex differences were found on achievement measures within or between classes. 2) A reading attitude measure showed more positive scores toward reading (from beginning to the end of the year) in both classes. The only difference between the two classes concerning content was that the class that had the high interest textbooks chose reading significantly more frequently as a free choice activity. 3) On a textbook evaluation questionnaire, both teachers rated the story content in the new primers as more interesting than the content in the previous series. Both classes had significantly higher reading achievement scores than the previous year's classes, with the same first grade teachers. It was the impression of the investigators (Zimet et al., 1974) that the experimental conditions were responsible for the equal reading achievement of the two classes and their higher achievements as compared to previous classes.

Evidence for the influence of content on first grade children has also been accumulated through less global investigations.

Between 1962 and 1973 (Blom et al., 1968; Zimet et al., 1973), a number of published and unpublished studies focused on children's responses to content.

From reading a large number of commonly used primer textbooks published before 1962, these investigators developed three impressions: 1) that the content presented more girl activity than boy activity themes; 2) that boy activity stories ended in failure more often than did girl activity stories; 3) that the activities were of interest to children younger than first grade. A questionnaire, containing a sampling of primer stories with ratings on these three dimensions, was presented to 150 college freshman English students and the results confirmed the impressions of the investigators. Pictorial rating scales were then constructed on the dimensions of sex of activity, age of activity, and outcome of activity. A number of stories were read to children individually in two first grade classes from a middle class suburban school. Again, the children agreed with adult ratings of the same stories, except for outcome where the stories had laughing and smiling characters associated with failure—an ambiguous situation indeed. However, when the pictorial scale study was repeated with a group of migrant first grade children, the ratings were very different. Inquiry revealed that the activities in the stories were, in general, foreign and unknown to them. As a consequence, the activities were rated older because they might occur later in the children's lives (Blom, 1975).

Stories were also obtained by trained nonwhite interviewers from Chicano and black children of differing socioeconomic status. The children were asked to tell stories that they would like to read in school and these stories were tape recorded. In review, the content of these tapes contained far more lively fantasy, real life drama, aggressive action, and affects than that displayed in first grade reading textbooks.

In another study (Wiberg and Trost, 1970), a comparison was made between the content of first grade primers and the content of library selections made by 45 first grade students during an academic year. The same content analysis scheme previously referred to was used (Blom et al., 1970). Three populations of stories were compared: a national sample of primers, the checkout library sample, and the noncheckout group. As might have been expected, the library sample available for first grade readers differed markedly from the national sample as to content. The typical selected library story tended to cluster into two groups: 1) fun and fantasy promoting, with highly creative imagery, plot, and characterization and 2) adventuresome, appropriate for age level or relevant to a more advanced age, informative, and life oriented realism. A gross distinction was evident between what children want to read (selected library stories) and what children have to read (national primer sample).

This study was replicated by Zimet and Camp (1973) on 152 first grade students in an innercity school over a five month period. The fourteen most preferred books of the suburban school population were *similar* in most content characteristics to the most preferred list in the innercity school population.

It was also demonstrated that first grade students had preferences for content of stories when a paired comparison choice was presented to them (Rose et al., 1972). Five pairs of stories were selected from a pool of 117 primary level stories chosen from 1,307 stories in the national sample of primers. Each story in a pair was contrasted on one content variable only, such as pranks compared with a Pollyanna theme. The other paired stories dealt with age of activity, character interaction, sex of activity, and outcome of activity. In each paired comparison, one story was judged to be developmentally relevant and the other inappropriate. Significant differences in preference that favored relevancy were found for theme, character interaction, and sex of activities for a group of seventy-six first grade children. While these content variables were selected from commonly used first grade primary stories, in reality there was a dearth of such content in them.

In summary, empirical findings, clinical and impressionistic reports, global studies, and indirect evidence from more limited, focused, systematic investigations further substantiate that content of reading material has a motivational influence on learning to read. There is evidence that preferences, reading achievement measures, interest, enthusiasm, reading use, awareness of meaning, and relevance exist in children in relation to content. Content should be thought of broadly to include themes, language, and words which are developmentally, culturally, socially, and naturalistically appropriate to children.

Cross national studies

Another way to obtain perspective on one's own country is to make comparisons with other countries on the same dimension of study. While the motivational dimensions of content of primers previously mentioned are somewhat culture bound, it is possible to assess attitudes in the content of stories from different countries. This has been done in a number of studies (Holmes, 1960; McClellan, 1961; Klineberg, 1963).

An attitude can be defined as a personal or impersonal attribute to which value can be assigned such as good-bad, desirable-undesirable, or important-unimportant. Stories present attitudes and, from the standpoint of textbooks, present a means of socialization for a given country. Therefore, content is important in developing attitudes which are of significance, in their own right, and also interact with content as motivation. For example, if a country displays attitudes of learning and schooling as important

in textbooks and at the same time has interesting, appropriate, and relevant information in texts, then attitudes and motivation have a facilitating interaction.

Furthermore, when a country provides each first grade child his own reading book to keep and use as she/he wants, this presents a different communication than when a country provides an already used book to be borrowed, left at school, and considered school property. The practice of a child borrowing or loaning a reading book was one of 162 scorable questions asked reading specialists in twelve countries (Denmark, England, Finland, France, Greece, Iceland, Ireland, Norway, Scotland, Sweden, United States, and West Germany). This questionnaire study (Blom et al., 1976) found that among these countries only England and the United States loaned books. The study further analyzed the results on questionnaire items which were grouped into six factors: language, family, teacher, country, school, and pupil.

Some unpublished results from this study are of particular interest in relation to content of textbooks. Except for France, all of the twelve countries relied considerably on textbooks for reading instruction. The authors of these textbooks were primarily women in three countries, primarily men in three, and both men and women in five. In one country, there was no agreement among the respondents about authors. Respondents were also asked to rate the appropriateness of content of introduction to reading textbooks in their countries on eight dimensions. The results are reported in Table 1 in terms of all the countries and the United States.

The findings in Table 1 indicate that some respondents did not answer some of the questions or were not in agreement with another respondent from the same country, suggesting a lack of clear impression (No Reliable Date - NRD). The United States does not compare favorably to the other eleven countries as a group on any content dimension. When these same results are viewed country by country in Table 2, the Inappropriate (IA) and Appropriate (A) ratings become clearer.

The data from Table 2 suggest that some countries give more attention to content and others do not. England and United States appear not to do so. This is not just a reflection of English speaking countries since primers in Scotland are viewed as having appropriate content.

While one can question the reliability and validity of questionnaire data findings, the authors of this study would point out that opinions from educators knowledgeable in reading are significant, that there are limited international data published on many questions of importance to reading, and that at least the results have considerable heuristic value. Unfortunately, studies on reading tend to be natiocentric and, as Jansen et al. (1978) have stated, reading and reading research may be viewed primarily as an

Table 1
Ratings of Respondents from Twelve Countries in Relation to Primer Content

Respondent Ratings	Primer Content Dimensions: Overall	Ethnicity	Nationalities	Regions of Country	Girl Interest	Boy Interest	Age Appropriate	Real Life
Appropriate (A)	5	4	3	6	6	8	5	3
Inappropriate (IA)	4	4	4	4	2	3	2	3
No Reliable Data (NRD)	3	4	5	2	4	1	5	4
USA Results	IA	IA	IA	IA	NRD	IA	NRD	IA

Table 2
Primer Content Ratings by Respondents from Twelve Countries

Country	Rated Appropriate	Rated Inappropriate	No Reliable Data
France	0	8	0
England	1	7	0
United States	0	6	2
Norway	0	1	7
Iceland	1	0	7
West Germany	2	1	5
Finland	3	0	5
Ireland	5	1	2
Greece	6	2	0
Denmark	7	1	0
Sweden	8	0	0
Scotland	8	0	0

Anglo-American activity. Downing (1973) has indicated the numerous contributions which come from comparative studies and has published findings from thirteen countries of the world.

In an unpublished study, Blom (1972) examined both motivational and attitude contents in sixty stratified randomly selected stories from primers in six countries. A particularly interesting finding was that antisocial attitudes (physical aggressiveness, uncaring, selfishness, nonconformity, disobedience, and cruelty) were displayed almost exclusively by animal characters (pets) in the American primers and to some extent in English primers. In Russia, South Korea, Italy, and West Germany, these attitudes were portrayed primarily by children and adults. A provocative and disputable interpretation of this finding is that the United States contents displace and project antisocial attitudes rather than attempt to directly socialize them within cultural institutions.

National gestalts based on the content analysis data were formulated for each of the six countries. In the United States, textbook stories present the instrumental task of reading as an end in itself rather than a means to an end of acquiring information. The contents emphasize child and human life as pleasurable, caring, and nurturing, rich in imagination and regression, and ending happily. The primary institution of socialization is the family, and getting along with others receives maximum attention. Antisocial elements are not attributed to humans but are displaced onto animals, mainly pets. There is a lack of traditionalism and conformity. While intelligence, as well as incompetence, is stressed, males fail and need help to a great extent. The symbols of the culture are not highly presented. Frustration is high.

The potentialities of comparative studies on reading have been illustrated by the studies mentioned in this paper. A method of cross national comparison of the content of first grade reading textbooks has been described, where the use of descriptive and mathematical statistics is possible. As a result, one can identify similarities and differences between countries and relate them to findings from other investigations employing different materials and methods.However, it is difficult to demonstrate the direct influence on the reading process and the teaching of reading. Such influences can only be indirectly, although strongly, implied.

Summary

The content of reading textbooks and other books employed in the teaching and practice of reading can be considered to have a number of overall functions: 1) instrumental, 2) motivating, 3) socializing, and 4) therapeutic. These functions obviously interact. The degree of success in reading acquisition, achievement, maintenance, and use may well depend upon the content of books as motivating, socializing, and therapeutic agents. Studies of the content of American primers, responses of children to content, and comparative (cross national) reading suggest the importance of realism, relevance, interest level, functionalism, and purpose in the reading process for children.

Reading can serve multiple purposes: fantasy expression and gratification, channeling impulse expression, providing information about the world and fostering reality testing, presenting ideals and values to guide behavior, dealing with universal conflicts and feelings, nurturing cognitive development and functioning, substituting for motor action, liberating one's self from one's own environment, offering comic relief, obtaining pleasure in words and ideas, and preparing for and integrating life experiences. Content of textbooks should offer some of these potentialities and instruction should make early and greater use of reading materials freely

selected by children. Zimet and Blom offer a number of recommendations based on some of these purposes. The position is taken by this author that no matter how reading is taught, and whatever theoretical model for reading is used, the *why* of reading must always be considered for all children.

References

Allen, Van S. "An Analysis of Textbooks Relative to the Treatment of Black Americans," *Journal of Negro Education*, 40 (1971), 140-145.

Ashton-Warner, S. *Teacher*. New York: Simon and Schuster, 1963.

Blom, G.E., et al. "Content Analysis and Children's Responses to First Grade Readers in the United States and Foreign Countries," *Final Report, Project No. 5-0556*, Contract No. OE 6-10-123, Office of Education, Bureau of Research, U.S. Department of Health, Education, and Welfare, October 15, 1968.

Blom, G.E. "An Examination of the Content of Introductory Reading Textbook Materials in Six Countries," paper presented at the Annual Convention of the International Reading Association, Detroit, 1972.

Blom, G.E. "A Case Study of a Research Project on the Content of First Grade Reading Textbooks," in J.E. Anthony (Ed.), *Explorations in Child Psychiatry*. New York: Plenum, 1975.

Blom, G.E., M. Jansen, and P. Allerup. "A Cross National Study of Factors Related to Reading Achievement and Reading Disability,"in J.E. Merritt (Ed.), *New Horizons in Reading*. Proceedings of the Fifth IRA World Congress on Reading, Vienna, 1974. Newark, Delaware: International Reading Association, 1976.

Britton, G.E. "Sex Stereotyping and Career Roles," *Journal of Reading*, 17 (1973), 140-148.

Child, I.L., E.H. Potter, and E.M. Levine. "Children's Textbooks and Personality Development: An Exploration in the Social Psychology of Education," *Psychological Monographs*, 60 (1946).

Costo, R. (Ed.). *Textbooks and the American Indian*. San Francisco: The Indian Historical Press, 1970.

Davies, A. *Printed Media and The Reader*. Buckinghamshire: The Open University Press. 1973.

Downing, J. (Ed.). *Comparative Reading*. New York: Macmillan, 1973.

Fader, D., and E. McNeil. *Hooked on Books: Program and Proof*. New York: Berkley, 1968.

Fernald, G.M. *Remedial Techniques in Basic School Subjects*. New York: McGraw-Hill, 1943.

Frasher, R., and A. Walker. "Sex Roles in Early Reading Textbooks," *Reading Teacher*, 25 (1972), 741-749.

Gibson, E.J., and H. Levin. *The Psychology of Reading*. Cambridge: MIT Press, 1975.

Gough, P. "Sexism in Basal Readers: A Continuing Trend," *Indiana Reading Quarterly*, 6 (1974), 27-30.

Graebner, D.B. "A Decade of Sexism in Readers," *Reading Teacher*, 26 (1972), 52-58.

Holmes, M."A Cross Cultural Study of the Relationship between Values and Modal Conscience," in W. Muensterberger and S. Axelrad (Eds.), *The Psychoanalytic Study of Society*, Volume I. New York: International Universities Press, 1960.

Jansen, M., B. Jacobsen, and P.E. Jensen. *The Teaching of Reading without Really Any Method: Reading Instruction in Denmark*. 1978, in press.

Kane, M.B. *Minorities in Textbooks; A Study of Their Treatment in Social Studies Texts*. Chicago: Quadrangle Books, 1970.

Key, M.R. "The Role of Male and Female in Children's Books: Dispelling All Doubt,"

Wilson Library Bulletin, 46 (1971), 167-176.
Klineberg, O. "Life is Fun in a Smiling, Fair Skinned World," *Saturday Review*, February 16, 1963, 75-77.
Laubach, F.C., and Laubach, R. *Toward World Literacy: The Each One Teach One Way*. Syracuse: University of Syracuse Press, 1960.
McClellan, D. *The Achieving Society*. Princeton, New Jersey: Van Nostrand, 1961.
Murphy, H.A., et al. "The Spontaneous Speaking Vocabulary of Children in Primary Grades," *Journal of Education*, 140 (1957), 2-105.
National Organization of Women. *Dick and Jane as Victims: Sex Stereotyping in Children's Readers*. Princeton, New Jersey, 1972.
Roberts, J., and J.I. Baird. *Parent Ratings of Behavioral Patterns of Children in the United States*. DHEW Publication 72-1010. Rockville, Maryland: National Center for Health Statistics, 1971.
Rose, C., S.G. Zimet, and G.E. Blom. "Content Counts: Children Have Preferences in Reading Textbook Stories," *Elementary English*, 49 (1972), 14-19.
Stanchfield, J.M. *Sex Differences in Learning to Read*. Bloomington: Phi Delta Kappa Education Foundation, 1973.
Tittle, C.K. "Women and Educational Testing," *Phi Delta Kappan*, 55 (1973), 118-119.
Waite, K.R. "Black and White Families in a Multiethnic Urban Primer," in S.G. Zimet (Ed.), *What Children Read in School*. New York: Grune and Stratton, 1972.
Whipple, G. *Appraisal of the City Schools' Reading Program*. Detroit: Detroit Public Schools, Division of Improvement of Instruction, Language Education Department, 1963.
Whipple, G. "Multicultural Primers for Today's Children," *Education Digest*, 29 (1964), 26-29.
Wiberg, J.L., and M. Trost. "A Comparison between the Content of First Grade Primers and the Free Choice Library Selections Made by First Grade Students," *Elementary English*, 47 (1970), 792-798.
Zimet, S.G. (Ed.). *What Children Read in School*. New York: Grune and Stratton, 1972.
Zimet, S.G., et. al. "Influence of Content of Primers on School Children," *Final Report Grant 5 RO1 HO O3775 DBR*, National Institute of Child Health and Human Development, U.S. Department of Health Education and Welfare, July 31, 1973.
Zimet, S.G., and B. W. Camp. "Favorite Books of First Graders from City and Suburb," *Elementary School Journal*, 75 (1974), 191-196.
Zimet, S.G., et al. "Learning to Read and Liking It: The Relationship between Textbook Content and Attitude Toward Reading," unpublished report, 1974.

The authors of this article are critical of many assessment procedures used in schools. Because of the recognized importance of assessment, they have addressed the problem of how assessment can be tailored to fit the needs of the classroom. From your knowledge of school testing programs, would you agree with the authors' criticisms of school testing? Are there other misuses of tests not covered by the authors? In your opinion, have the authors suggested reasonable approaches to assessment so as to overcome the abuses?

What research can tell the reading teacher about assessment[1]

ROBERT C. CALFEE
Stanford University

PRISCILLA A. DRUM
University of California at Santa Barbara

RICHARD D. ARNOLD
Purdue University

How can assessment be tailored to fit the needs of the classroom? To answer this question, we need to consider three other questions:

1. Assessment for what? (The goals)
2. How to assess? (The methods)
3. Is assessment doing its job? (The criteria)

We will consider each question in turn as it relates to reading, but first, two major conclusions:

1. Teachers need to learn more about the process of assessment in order to assess for instructional purposes.
2. Classroom assessment ought to aim toward the precise and efficient measurement of specific component skills for short-term decisions.

The bulk of this paper will buttress and illustrate these two generalizations, but some preliminary comments will help to set the stage. Much research on reading assessment aims toward goals quite different from those that are foremost for the classroom teacher. The goal of conventional test construction (and thus of the research that centers around such test construction) is the production of reliable and substantial individual differences, based on stable scores for each student which place him at some point below, at, or above the average for some larger population. The aim

is an instrument for making major, long-term decisions (Carver, 1974).

The teacher needs information of much more immediate character. How well can the student read now? What specific reading instruction should he receive next? Is the instruction successful? To the question, "What does the research literature on assessment have to say to the classroom teacher about his instructional needs?" the answer is, "Not much!"

To our knowledge, no existing assessment system handles the range of assessment tasks encountered by the reading teacher. Most commercial tests provide little evidence for directing instruction, and are too expensive in time and effort to meet the teacher's needs. It is not that commercial tests are faulty, rather that they are designed for other purposes than immediate instructional decisions.

Where is the classroom teacher to find assessment tools that permit reasonable and sound instructional decisions? We believe that this is the wrong question. The problem of improving classroom assessment is not likely to be solved by the development of the "right" testing system. Public and private contractors may well improve the offerings available to the teacher. But we suspect that the answer to the task of classroom assessment lies in the teacher's knowledge of suitable processes to carry out continuous assessment. If so, the chief task of those who would improve assessment of reading for purposes of instruction lies not in psychometrics but in teacher training, where the goal of the training is to help the teacher to select with care from what is available for other purposes and to modify and simplify the materials at his disposal with an eye to practical application.

Goals of assessment

First, certain goals of assessment aim toward long term prediction. This is true in "Evaluation of the Individual" (Cronbach and Gleser, 1965), for job placement, for school admission, for a grade or achievement mark of some sort. It is true when assessment serves for evaluation of a program. The administrator has to decide whether a curriculum is effective, whether a special program is better than the regular program, whether extra money is making a difference. Diagnosis also falls in this category. Diagnosis is for special cases, like physical anomalies. A person who can't see well has trouble learning to read. If he can't hear well, he may also have trouble in school tasks, including reading. These are special cases and may require a clinical specialist.

Other goals aim toward short-term decisions. Assessment can guide the classroom teacher in instructional decision-making. The instructor has to stay current on what each student knows if instruction is to be precisely directed toward specific needs. *Individualization* is

the usual label for this concept. Each student is assessed as to his present skills, abilities, and knowledge, so he can be helped to move from where he is toward some reasonable goal.

In summary, assessment for short-term instructional decisions covers diverse situations: optimizing instructional sequences, measuring immediate response to instruction, regrouping for instruction for specific purposes, and deciding on selection and allocation of resources (who needs the aide's time, the tutor's time, the terminal's time?).

Present methods of assessment in education

Psychometrically "sound" tests in use today include normative and criterion-referenced tests; these two types of tests differ little in content, though designed for different applications (Green, 1976). A norm-referenced test shows how the student's score or the class's score compares with the other students or classes who provided the standards for the test. A criterion-referenced test provides a score for a student or a class based on the number of items mastered (answered correctly) compared with some absolute standard. Neither type of score tells the teacher what a student knows or does not know; direction for further instruction is not indicated. Both types of tests are standardized; an exact procedure for administration is called for, with little room for clinical probing. Most tests are group-administered and use a multiple-choice format to facilitate machine scoring. The content resembles goulash; though a subtest structure is often imposed on the test items, the high interest correlations belie the different names assigned to subtests.

It can also be said that these tests are reliable, that the student's relative standing is stable over time, and that they are highly predictive of one another (Bloom, 1964). They are time consuming to administer and generally are not capable of repeated administration—two or three times a year at most. They yield a single type of measure (percentage correct or some transformation thereof).

Such tests have been developed to meet certain implicit and explicit criteria. It, therefore, makes sense to consider the standards and criteria that apply to the construction, administration, and interpretation of a test.

Test standards

We want to examine briefly several criteria for evaluating tests: reliability, validity, appropriateness, independence, discriminability, cost, and repeatability. The first two usually are discussed in texts on testing and generally the others are not (e.g., Anastasi, 1968; Cronbach, 1970).

Reliability: Does the instrument provide a consistent measure?

In general, reliability refers to the degree to which a measurement is consistent. We can consider the consistency in performance when a person is tested with one form of a test and then retested with a slightly varied form. Several things have changed. The exact form and content of the test has changed. The student has probably changed. He may have learned something, he may have forgotten something, he may have a headache that he didn't have earlier. All these sources of variability tend to reduce the reliability in test-retest situations.

Test developers tend to emphasize *within-test* reliability. There are several ways of thinking about this form of consistency (Cronbach, 1970, Ch. 6). For instance, suppose you divide the items in two at random and correlate the two subscores. Repeat this operation for all possible split-half divisions of the test, then compute the average correlation (Cronbach, 1951). This provides a measure of the extent to which each item contributes consistently to the total test score. One way to obtain "perfect" intratest reliability is to use a test in which the student either fails all items or passes all items. Test developers, to the degree that they strive for intratest reliability are under pressure to eliminate test items that yield divergent patterns of performance from one student to the next. The items that remain seem likely to measure general intellectual characteristics rather than performances that reflect specific instructional outcomes. So, if you want a perfectly reliable test, ask the same question twenty times. Either a student knows the answer or he doesn't. This would be absurd, of course, but in the limit it is the "ideal" toward which reliability aims.

Intratest reliability is important when the test score is to serve for a major decision, but it may be counterproductive for instructional decision-making. Teachers need to know more than the student's general ability. Individualization requires knowledge of diverse patterns of performance on specific tasks for different students. For the teacher, a *reliable* assessment instrument is more properly defined as one which accurately and consistently indicates the specific patterns of instruction that best fit the student's needs and capabilities (Calfee, 1976; Calfee and Drum in press).

Validity: Does the instrument measure what it's supposed to?

As with reliability, the concept of validity assumes many guises. *Face validity* means that the test looks like it measures what ought to be tested. *Construct validity* means that if several tests seem to be measuring the same thing, there must be something there to be measured. *Predictive validity* means that there is a correlation

between a test and a criterion of performance (usually another test).

To possess adequate validity, a test usually has to satisfy each of these criteria. For instance, one can predict reading achievement reasonably well from mathematics tests, but teachers and parents would question the face validity—it would not be seemly to measure reading performance with a test containing arithmetic "word problems," even if the test met the usual standards of reliability and predictive validity.

Appropriateness: Does the instrument measure sensibly, given the way the evidence is to be used?

Appropriateness is introduced here as a fuzzy concept covering several related matters. In part, it has to do with whether a test is linked to the goals of an instructional program with sufficient directness and breadth. Researchers learn the meaning of this concept when public school teachers ask why no one tests what they teach. This complaint is fair and deserves the attention of evaluators.

Appropriateness is disregarded in the common practice of assigning a student to a particular level of a test according to his age or nominal grade placement rather than his actual performance level. The experience of the Chicago schools when they selected achievement tests to be appropriate to the students' actual reading level is most instructive in this connection (Figure 1). Asking a high school student who reads at the first or second grade to handle an advanced level of the Metropolitan Achievement Test is a mistake; whatever the score, it is unlikely to reveal the student's actual skill in reading. The advanced test is designed for those reading at grade six or above. Students reading at a lower level are likely to guess randomly at the answers, but this performance is likely to lead to a grade level score that is higher than their actual reading ability.

Independence: If several skills are measured, is there evidence that they are more or less separable and autonomous—not closely correlated?

To be most useful, the several scores from an assessment battery should provide the teacher with distinctive pieces of information. When all the subtest scores are highly intercorrelated, the teacher receives little guidance about distinctive courses of action. As Thorndike (1973) has pointed out, even a modest degree of correlation between two scores makes it difficult to make differential diagnosis, given that the scores are normally distributed. The magnitude of this problem for certain commercial tests has been discussed by Calfee and Venezky (1969), and possible remedies

New testing method cited for drop in reading scores at public schools

By Casey Banas
Education editor

A NEW READING test report of Chicago public elementary schools showed Wednesday that the scores are down, but school officials say it's not because children are performing any worse than last year.

John Wick, board of education director of research and evaluation, said the scores are lower because of a new testing method.

In past years, many children took tests too difficult for them and the results were no higher than if they merely guessed at answers in the multiple choice exams. This year, the pupils were given tests at levels that more accurately reflected their actual reading ability.

"THERE IS NOTHING to suggest the children are reading more poorly," Wick said. He emphasized the results give a more accurate picture of actual reading levels.

Scores of some schools dropped as much as 1.8 grades using the new testing method.

Unlike other years, the 1975 reading score report does not compare the entire system's average student to national norms. But it does show, for example, that the average 13-year-old pupil is reading at or above the national norm for eighth graders in only 40 of 359 elementary schools.

AND FOR THE first time, the report included a chart showing that the higher the percentage of children from poverty level families in the school, the lower the test scores will be. Mean scores for 8 and 12-year-old students varied two to three years between the highest and lowest poverty level schools.

Figure 1. Effect of test selection on reading scores (*Chicago Tribune*, December 4, 1975). Reprinted, courtesy of the Chicago Tribune.

suggested (Calfee, 1976). One desirable condition is that each test be "clean," i.e., that steps be taken to insure that the test measures the desired skills, and none other. Later, we will describe a second approach built upon factorial test design, in which systematic variation in the materials and conditions of testing allows the tester to find out the circumstances under which a student can and cannot handle a task.

Discriminability: When possible, information from a subtest should be "yes-no."

It takes more expertise and attention to monitor the ammeter in an automobile's electric system than to notice simply whether the

generator light is on or off. Similarly with a test—when the scores on a test take the form of a normal distribution, then fine gradations in performance matter a lot and interpretation is more difficult. It is much easier to interpret performance when it is either clearly at the mastery level or altogether faulty, with no "in between" scores. Careful specification of the task is required, but the benefits for instructional decision-making can be considerable (Calfee, 1976).

Cost: How much time and money to buy, administer, score, and interpret?

Tests cost money, and they cost time. These costs may be overlooked by teachers, even when they are the ones who pay. For instance, in one school, teachers spent three days testing the students' reading skills in third and fourth grades. The scores were then used for the sole purpose of sorting students into three reading groups: high, medium, and low. Obtaining a ten minute oral reading sample from each student would probably have done the sorting job as well or better, and at much less cost. When a major decision is to be made, substantial cost is justified; when continuous short-term decisions are required, low cost is essential.

Repeatability: For classroom instruction, it should be practical to readminister a test whenever the teacher needs information

The time and cost required by many tests makes repeated administration impractical. Besides this, the psychometric concern with reactivity in retesting leads to advice against repeated administration of the same form. It is rare to find more than two alternate forms of most commercial tests. For evaluation of a program or an individual, assessment once or twice a year is sufficient. But the teacher who wants evidence on the effectiveness of yesterday's instruction needs an "off-the-shelf" test, one which comes in many forms and can be used as often as necessary.

Standards for practical classroom assessment

Testing for different purposes demands attention to different criteria. For example, program evaluation requires high reliability on a small number of dimensions; one dimension may suffice. The cost is considerable for a reliable judgment, but it is justified when a major (perhaps irreversible) decision is to be made: Does this student go to college? Should this child be placed in a special education classroom?

As a different example, program evaluation can be accomplished effectively with short tests that are unreliable for each

individual student, because the focus is not on the individual but the program. Matrix sampling fits this need well; a variety of test items is sprinkled over many students to obtain the needed information at considerable reduction in cost to the schools, the state, and the teachers.

A cursory examination of the research literature reveals the emphasis on tests suitable to long-term, major decisions (e.g., Weintraub et al., 1974: 460-454; 1973: 429-447). The teacher's need for in-class assessment, on the other hand, is best met by tests that are speedy, precise, clearly appropriate, and flexibly repeatable. The research base for such tests is scanty, to say the least. The purpose of instructional testing is to pinpoint, at a given time, exactly *who* needs help and guidance in learning *what* topics. A great deal of time is lost in the school by teaching students what they already know, which means that resources are wasted and that students who need help are less likely to receive help.

We can highlight several desirable characteristics of tests for short-term instructional decisions. They are easily readministered in one form or another. They need not be absolutely standardized; rather, the teacher has room for clinical probing. Variations in testing procedure need not be random; the test can be quite systematic while permitting flexibility (e.g., Calfee and Hoover, 1974). The test must permit easy decisions by the teacher—waiting for the outcome of a stepwise, multiple, linear regression followed by discriminant analysis won't work for the teacher.

The teacher cannot expect to find on-the-shelf tests that are well suited to short-term instructional decisions. Moreover, training on "test construction" reflects the conventional psychometric tradition, and so the teacher is likely to be poorly prepared to select, to adapt, and to create useful instruments. It is not the intention of this paper to go into detail about the program of teacher training that might alleviate this gap. However, we suspect that it would center about an analytic approach to what is being taught—we have referred elsewhere to the distinction between a "Jello" model of the mind in contrast to the "works in a drawer" model, the former being more Gestalt-like, the latter more analytic and information processing in character (Calfee and Floyd, 1973).

A Jello model yields simple descriptors for students: *smart, low ability, disadvantaged, minimally brain damaged, mentally gifted.* The works in a drawer model describes a student as knowing certain things (e.g., letter-sound correspondences), but not knowing others (e.g., the principles for short and long vowel alternations). In our experience, many teachers tend toward the Jello model in thinking about assessment, and in their approach to instruction as well. By focusing on specific skills (not quite the same as behavioral objectives), the teacher thinks more precisely about what is being assessed, and what the suitable instructional response might be. In

this frame of mind, it is easier to develop a "clean" test which aims at a precisely defined skill or knowledge. Our research suggests that it is not unusual for such a test to have an all-or-none, "on" or "off" character (Calfee, 1976). For instance, Figure 2 shows the distribution of test scores when students' knowledge of vowel digraph correspondences is tested with care and precision.

An analytic frame of mind leads the teacher to take advantage of variation in the testing context, rather than restricting himself to a standardized set of conditions. It is relatively easy to demonstrate that a student will fail a task under a set of conditions; it takes more work and imagination to find those conditions under which success, however limited, is possible. We have proposed the application of factorial-design technology in the development of testing systems for exploring the limits of a student's knowledge of decoding principles (Calfee, Fisk, and Piontkowski, 1975). We believe the teacher can profit from learning the concept of factorial variation in test conditions as a way of exploring the workings of the student's head—first graders can be trained to benefit from application of factorial principles to problem solving (Anderson, 1965; Inhelder and Piaget, 1958: Ch. 3) and certainly the discovery of what a student knows about reading is a problem solving task par excellence. An important note: efficient application of complex test designs requires good preliminary screening. The tester needs to decide quickly where to focus his assessment activities.

Each of these issues—clean tests, factorial test design, and efficient entry testing—is a complex matter. We cannot do more than emphasize a few of the main points.

Clean tests

A clean test is one in which a single, well-defined component is examined (Calfee, 1976; Calfee, Chapman, and Venezky, 1972). The test begins as simply as possible; ideally, no student should make a mistake under the simplest conditions. This shows the student understands the nature of the test and can handle the general test-taking requirements. Then the difficulty of the test is increased systematically. As errors occur they indicate the nature of the student's problem. Developing a clean test often requires working backwards, asking the question, "What must the student know to be able to succeed in this task?" In answer to the question, "What does a failure mean?" the teacher must make a guess. Based on the guess, the teacher decides how to simplify the test. If the guess was correct and the student is now successful, his problem has been isolated. If he still makes mistakes, the guessing-testing process is pursued further.

The major barriers to a clean test are often the general test requirements. To do well on a test, the student must understand

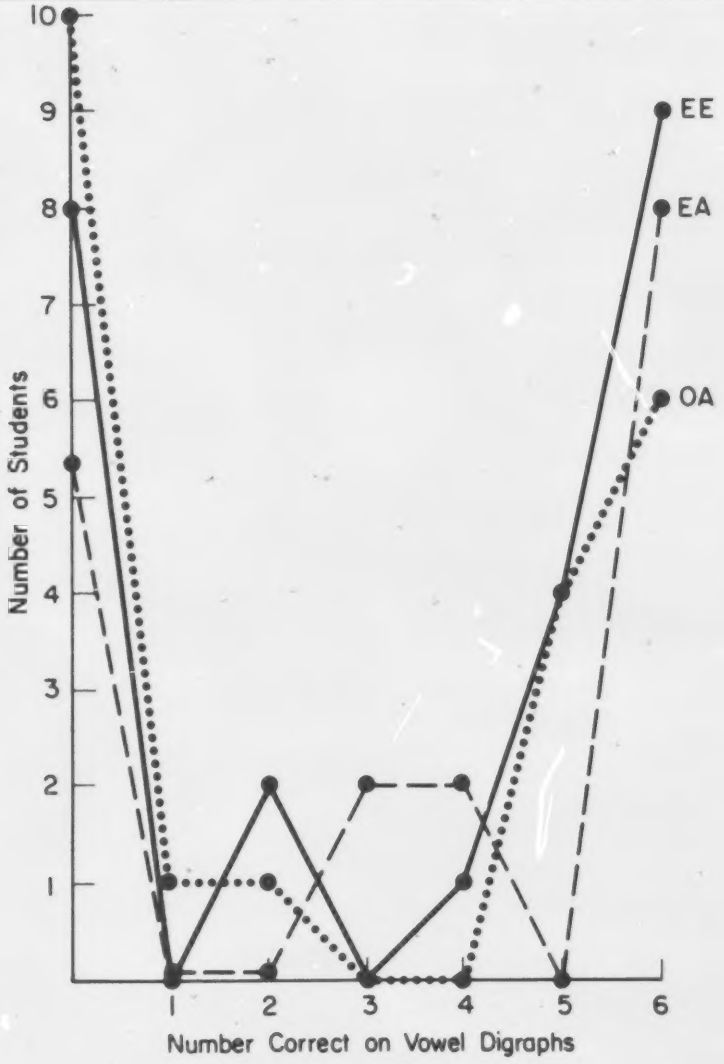


Figure 2. All-or-none distribution of student responses on test of knowledge of vowel digraphs (after Campbell, in preparation).

what is expected of him, and he must feel encouraged and motivated to do well. Listening carefully and following instructions are important for success, and some students are better than others at these general skills. Individual or small group testing makes it easy for the teacher to assure that all students know what they are to do, and makes it more likely that performance will reflect specific rather than general skills. The clinical tester receives the training needed to gain understanding. The classroom teacher may not have had any such training, but he can be aided by guidelines for determining readiness for a test and suggestions about how to promote readiness.

Factorial tests

Complementing the notion of a clean test is the idea of factorial test structure. The clean test approach aims toward constancy in all dimensions of the test except one; the factorial approach aims toward systematic variation in several dimensions of the test. Because the concept is new, we will illustrate in Figure 3 how a factorial structure provides a framework for the instructor to think about in testing reading comprehension. One dimension is the nature of the task: oral reading, silent reading with no time pressure, and silent reading with time pressure. As a student becomes competent he should be able to perform well and equally so under all these conditions.

A second dimension of the factorial test is the "question mode." How shall the teacher request information from the student after he has finished reading? Perhaps the simplest approach is to ask him direct, literal questions—these can be quite specific or may allow for a more general response to the passage. A recognition test is slightly more difficult, because the student has to read the question and the alternatives, but at least the answers are provided to him. Production and essay tasks demand even more from the student. To summarize a story requires some sophistication, and failure can be traced to any of several possibilities. If performance has been measured under simpler conditions, most of these possibilities can be evaluated. Variation in materials is the third major dimension. It makes quite a difference whether the student is reading a familiar or an unfamiliar topic; difficulty level of vocabulary also makes a difference.

Envision each student's performance in the multidimensional space of Figure 3. The task of the instructor is to locate the student in this space, in the sense that the instructor knows whether the student can perform accurately and quickly in each cell. In fact, one might conceive of testing that aims to trace through the three-dimensional space a line that represents the boundary between where the student can perform adequately and where he has trouble.

TASK		MATERIALS			
		Familiar Topic		Unfamiliar Topic	
Reading Mode	Question Mode	Easy Vocabulary	Difficult Vocabulary	Easy Vocabulary	Difficult Vocabulary
Read Aloud	Oral/ Literal				
	Recognition/ Literal				
	Recognition/ Interpretative				
	Production/ Essay				
Read Silently No Time Pressure	Oral/ Literal				
	Recognition/ Literal				
	Recognition/ Interpretative				
	Production/ Essay				
Read Silently Time Pressure	Oral/ Literal				
	Recognition/ Literal				
	Recognition/ Interpretative				
	Production Essay				

Figure 3. A factorial structure on dimensions of reading for instruction and assessment.

Entry level assessment

For reading, we have devised a fairly efficient and effective means to sort students (Calfee and Hoover, 1974): Choose a few lists of words arranged by difficulty level, then say to the student, "Here are some word lists I'd like you to read. Which list, A, B, C, D, or E, do you think you can read?" As soon as the student has pointed to the list he thinks he can read, the teacher has a piece of useful information. If the student's self-assessment agrees with his subsequent

performance, he knows realistically what he can do. If he performs two or three levels below his estimate, he at least has a good self-concept.

The teacher then asks the student to read the list he has just pointed to. If he has trouble with several words, the teacher asks him to try an easier list. If he pronounces every word quickly and correctly, the teacher asks him to read a harder list. The student will reach the limit of his skill within a few minutes. A similar procedure is used to assess the level of understanding of word meanings and of paragraph comprehension.

We have used a test built around this model for research activities, and are pleased with the rich return from what is generally less than a twenty minute test session. But the point to be stressed here is the value of this test for purposes of determining entry level to other tests (and to instruction, of course). Precise assessment of a student's skills and knowledge, if it is to also be efficient and not time consuming, requires a quick screening to determine relative standing in different component areas of reading.

Categories of reading skills

Reading includes several areas of knowledge and skill, and any analytic effort to assess "reading" must attempt a "first cut" of the collection into reasonably digestible pieces. This list will serve for the moment: decoding, vocabulary, grammar, transliteral comprehension, and inferential comprehension. Each area will be discussed in more detail. The reason for the division of reading into these components is straightforward: methods of assessment and instruction are distinctive for each component, in our opinion. The techniques for evaluating this hypothesis (for such it is) are also straightforward (Calfee and Elman, in press), although to the best of our knowledge, there is only a smattering of research that can be brought to bear on the issue.

Decoding

One set of skills centers on decoding, the translation from print to sound (Figure 4). It is not clear at what point during the acquisition of reading that the student can best develop this skill. But a good deal of data exist to support the proposition that the reader of English who can't look at new sets of words and decode them with fluency is likely to have trouble becoming a skilled reader. A person can learn something about reading without learning the skill of decoding, but it is unlikely that he will become a good reader.

Decoding instruction ends at about the third grade in the United States. Principles for translation of longer (more information-laden) words are covered badly (if at all) by most available pro-

grams. Teachers are ill-prepared to know when the instructor's manual contains nonsense. Instruments for direct assessment of decoding skills are rarely used in large-scale evaluation efforts—the multiple-choice, group-administered test does not lend itself to the measurement of pronunciation, which is the most appropriate way to measure the translation process. Middlegrade teachers, school administrators, and even the National Institute of Education have expressed concern about the problem of "comprehension," assuming more or less explicitly that the task of fluent decoding is well learned in the early grades (NIE, 1975: iv). Solid evidence on this point is scanty.[2] However, one can find data showing that 1) certain reading programs and approaches are less effective in promoting the acquisition of decoding skills (Barr, 1974; Chall, 1967), 2) many children still show lack of skill in handling basic decoding skills as late as the end of fifth grade (McDonald and Elias, 1975), and 3) substantial correlations are found between decoding skill and school performance up through college (Venezky, 1974b).

Vocabulary

Vocabulary development is recognized by most reading teachers as vitally important to success on academic tasks. Students need to be able to define words, synonyms, and recognize common usage of words (Figure 5). The science question in Figure 5 requires understanding of the word *orifice*. The dictionary definition is a start. But few words in English have single meanings, and the most common words have the largest number of meanings. Furthermore, even if the student were to internalize the dictionary, society and individuals keep devising idiosyncratic meanings.

DECODING = TRANSLATION FROM PRINT TO SOUND

PHRAIM	⟶	"F-R-$\bar{A}$-M"
COTHE	⟶	"K-$\bar{O}$-TH" (VOICED)

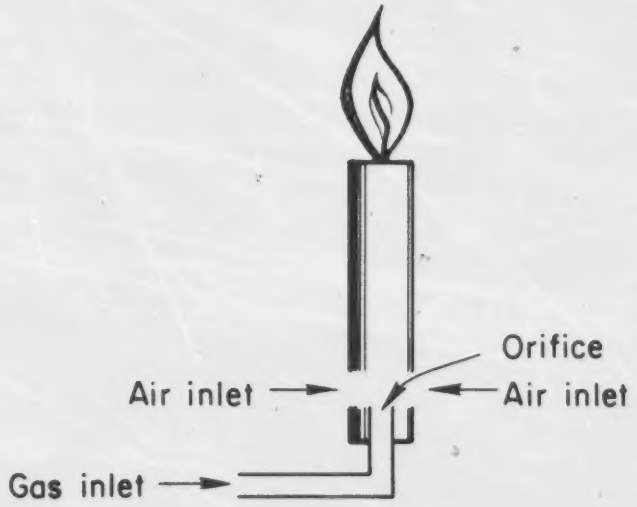
Figure 4. Examples used for assessing print-to-sound correspondences.

VOCABULARY = DEFINING, KNOWING, SYNONYMS, RECOGNIZING USAGE

WHAT IS AN ORIFICE?

"A MOUTH OR SIMILAR OPENING; A HOLE; AN APERTURE"

WHAT IS THE ORIFICE IN THIS PICTURE?



Typical gas burner

Figure 5. Sample science test item with illustrations.

Understanding of a word can be helped along by studying pictures, "vis ed" cards, or Language Master Cards, but adequate vocabulary is facilitated only as the student does a lot of reading and listening. The aim is to establish in the student's head a basic meaning for the word that meshes with a larger structural ensemble. A student knows the meaning of a word when he can understand it in various contexts and can recognize the common feature in a different context (e.g., he enlarged the orifice), or use it appropriately in speech and writing. An assessment task should consider not only whether the student knows a word, but at what level.

Grammar

Some may find it quaint to include grammar as part of the reading process, but it probably has as much place as comprehension skills. In both instances, understanding requires the transfer of skills from oral language to a new context, and the expansion and elaboration of those skills to meet the peculiar demands of the written language. Capitals, commas, periods are not directly indicated in speech. Sentence fragments are as common in conversation as are complete sentences. The physical and social contexts of speech disambiguate elliptic utterances. The reader, on the other hand, must note both punctuation and word order to obtain meaning, for he has no other guides. Olson (1975) has pointed out other distinctive differences between the written and oral version of the language.

When discussing grammar, an important distinction exists between style and substance. Style refers to following the proper convention: producing all the past and plural markers, using proper word order and the like. If the student is going to speak or write English "properly," he will have to know one or more sets of conventions, and be able to use them in the proper context. The conventions do not stand in the way of meaning most of the time. Some comedians make fun of English by convoluting the syntax, yet the meaning remains clear. To be sure, they would be marked down if an English teacher were grading them, or if the reading teacher were evaluating their oral reading (Lambert, 1972).

There are also substantive matters in grammar. Sometimes meaning is disambiguated only when the plural marker, the past marker or some other morphological ending is noted. If a particular word order has one meaning and a different word order conveys a different meaning, a substantive difference in grammar is apparent. "Bill told Jane to snitch the ice cream." has a different meaning from "Bill was told by Jane to snitch the ice cream." The answer to "Who will be punished for snitching the ice cream?" depends on recognizing this difference.

The whole problem of anaphora, of knowing that an indefinite reference is grammatically linked to certain places in a prose passage, is substantive. To read a paragraph and make sense of it, one must know enough about the grammar of the language to figure out these linkages. The passage in Figure 6 serves as an example. The questions about the story can be untangled only from knowledge of English syntax. Many children come to school with adequate knowledge in this area; others may need some help. It is the task of instructional assessment to distinguish one group from the other.

GRAMMAR = UNDERSTANDING SENTENTIAL AND TEXTUAL RELATIONS FROM GRAMMATICAL CUES

MARY SPRINKLED JOE. BECAUSE IT WAS A CHILLY DAY, HE WILL PROBABLY GET A COLD.

WHO GOT WET? WHEN IS THE ACTION TAKING PLACE? WHAT DOES "IT" REFER TO? WHO IS GOING TO GET A COLD?

Figure 6. Sample passage and questions illustrating referential complexity.

Transliteral comprehension

Comprehension is a complicated matter; it can be virtually synonymous with thinking. Trying to analyze the process of comprehension is an interesting challenge. We propose here two broad categories of comprehension tasks, transliteral and inferential.

Transliteral comprehension requires the student to have meanings for the words, recognition of word order, and either direct or analogical experience with the content, so he can extract and remember information conveyed directly by the passage, information fairly close to the surface (Figure 7). In this example, it takes only modest understanding of the paragraph to answer the question correctly. The paragraph links *high* by hyphenation with *crowned*,

TRANSLITERAL COMPREHENSION = USING SYNTACTIC, VOCABULARY AND TEXTUAL CUES TO EXTRACT AND REMEMBER INFORMATION CONVEYED BY THE TEXT.

A wide-brimmed, high-crowned Mexican sombrero, or hat, can have several uses. You can carry water in its deep crown. Pull it down over your face, and you can nap in bright sunlight. You can carry bananas in its sturdy brim. When the straw finally wears out, you can feed the sombrero to a burro.

A. The paragraph says a sombrero has a high:

1. brim
2. crown
3. decoration

Figure 7. Sample passage and a transliteral comprehension test item.

so the reader can couple *high* with *crown*, an easy matching task. If the target sentence were "The wide-brimmed Mexican sombrero, with its tall crown, can have several uses," the necessary information for transliteral comprehension is still available, but more understanding is needed. It is easy to imagine other ways of modifying the sentence so that the key information is still available, but less readily so.

In transliteral comprehension, all of the information needed to answer the questions is contained in the passage. Some questions can be answered by using matching techniques, some by prior experience without reading the passage, and some require an understanding of key terms. Useful assessment procedures sort out the strategies used by students to answer various types of questions.

Inferential comprehension

There is a kind of comprehension that requires a broader and

INFERENTIAL COMPREHENSION = USING INFORMATION FROM TRANS-LITERAL COMPREHENSION, PLUS EXPERIENTIAL KNOWLEDGE PLUS REASONING, TO GO BEYOND THE INFORMATION GIVEN.

The Everglades, a swamp area covering over 5,000 square miles in Florida, is a storybook jungle.

Numerous channels wind through saw grass and gama grass marshes. Rare fish and alligators live in these murky waters. Some waterways are tunnels walled by tangled roots of mangrove trees. On the roots creep brilliantly banded snails. Scorpions and cottonmouths slither among the logs.

Cypress and palmetto trees form a dense growth, where exotic tropical birds make their nests. Among these are the ibis, egret, and jacana. Biting, buzzing insects fill the air.

As you might expect, few people live in this wild but beautiful region.

This area is very appealing to

1. story writers
2. naturalists
3. swimmers

Figure 8. Sample passage and an inferential comprehension test item.

deeper analysis of the textual information (Figure 8). For instance, what must the reader understand in order to know that the answer in Figure 8 is probably "naturalist?" If the student isn't familiar with the Everglades, he may be taken with beautiful waterways, a thick jungle, scorpions (who knows what a scorpion is?), cottonmouths—it may sound rather charming, a lovely place to go swimming! (If the reader consults the American Heritage Dictionary, the only clue that all is not idyllic is the note that the scorpion has a "venomous" tail.) This passage describes a lovely place where one can enjoy a nice quiet swim. For those who are more informed, background and experience tell us that one doesn't swim in the wilds of the Everglades; the wise person stays in the airboat. One might choose "storywriters" as the answer, for the first sentence of the passage ends with "a storybook jungle." This "wild but beautiful" setting is perfect for a mystery or a romance. The preponderance of information on vegetation and rare wild life is the only clue that the test developer meant to focus on a naturalist's interests.

Here is another typical "comprehension" question:

"Most of the women in the United States are _____ .
a) plumbers, b) citizens, c) redheads, or d) waitresses."

With no passage to read, how does the student select the right answer? The task is only modestly related to reading, though it comes from an actual comprehension test. The student unfamiliar with our culture might think that "redheads" is right; "waitresses" makes sense if many of the women in his experience have been waitresses. An advocate of the women's liberation movement might choose "plumbers." The "correct" answer to the question actually seems stilted and perhaps absurd. The student must rely on knowledge and experience that goes beyond the question and looks at the demand of the t .sk. The good reader brings to bear on the topic what he knows, what he learns from the passage, and what he can figure out about the tester's reasoning and intentions. The teacher needs to know which of these is behind the "poor" student's failure.

Many students know that the right strategy is generally to read the questions, answer the ones that they already know, then go back and read the text to answer the rest. This strategy gives high payoff. For instance, consider the two questions in Figure 9. Both of these questions were accompanied by passages in the original test. In one case, astute "readers" can immediately deduce what the passage must have said; this analysis of the questions is part of what is meant by inferential comprehension. The other question requires detailed examination of the passage. The knowledgeable reader will immediately detect from the form and substance of the questions that he must refer back to the passage if he did not read it in detail to begin with.

Which of the following pairs of kings were father and son?

_____ (a) Alfred and Svein _____ (c) Svein and Cnut
_____ (b) Olaf and Svein _____ (d) Alfred and Cnut

One's recognition vocabulary is

_____(a) the same thing as his active vocabulary
_____(b) less often used than his active vocabulary
_____(c) usually smaller than his active vocabulary
_____(d) much larger than his active vocabulary

Figure 9. Two comprehension test items illustrating role of background knowledge.

The teacher who wishes to "measure comprehension" should be prepared to cover the full range of the student's skills. These include not only finding facts and making simple inferences, but also solving the problem of when to do one or the other. Moreover, the making of inferences is not only a logical process. Many comprehension questions require a process of inference that is more analogical than logical. This requirement seems altogether reasonable because life experiences are often based more on metaphor than logic. We make comparison with experience and fill in the missing parts of an event by analogy rather than by Aristotelian inference.

Task requirements in assessment

Five major areas of reading have been mentioned briefly: decoding, vocabulary, grammar, transliteral comprehension, and inferential comprehension. In examining these categories of reading skills, we also need to analyze the task requirements for successfull performance on a particular test within a given category. Some task requirements are specific to a given area, but others cut across all areas. For instance, the same basic situation may be presented to the reader so that he must recognize the correct answer from a set of alternatives, or must *produce* the correct answer from memory. The person's skill may allow him to perform well on one form of the task and not on the other. As Kintsch (1970: Ch. 5) notes, different performances under the two task formats permit the researcher (or tester) to infer underlying processes. Recognition of previously studied information suggests the information has been stored adequately; recall suggests that it was stored in a retrievable format.

To find out what a person really knows, the teacher must devise various ways to tap that knowledge. As noted earlier, it is relatively easy to show that the student cannot remember it under certain conditions. The most direct way to assess a person's knowledge is to ask him a direct question. If he doesn't give the answer, then a second, more probing question can follow. "Do you think it's this?" Maybe the probe will trip the memory key so the student responds with the correct answer to the original question.

Speed and *accuracy* comprise another important task dimension. Speed is not always "good," but often it is. Automaticity in basic skills can be especially critical (LaBerge and Samuels, 1974). For example, a few years ago we worked with some researchers who were developing a reading series for kindergartners. They had devised an algorithm for teaching children to decode. First the student learned a few letter-sound correspondences, then he moved his finger from one letter to another to blend the sounds: *b; b-a, ba; ba-t, bat.* Within a short time the kindergartners could decode a fairly substantial set of words. Some students were much faster than others,

of course. Some could look at the word and say "bat" and others were still going *b-a-t, bat*. Then they were asked to read sentences for the first time. The task changed from decoding one word at a time at a relatively easy pace to decoding a whole string of words. Furthermore, the children were expected to answer questions when they finished the sentence. A few seemed to become "instantly dyslexic" at this juncture in the program. In our opinion, this resulted from differences in speed of decoding. Speed of reading single words was not important *per se*. But it took so long for some students to translate the sentence word by word, that by the time they reached the end of the sentence they had forgotten the beginning. Since the decoding strategy didn't work, some students began to guess from initial letters, or they looked at the pictures, searching for meaning with little regard for the print—strategies typical of poor readers.

What is the import of the speed-accuracy distinction for the classroom teacher? Formerly, teachers were encouraged to test for both speed and power. Today, in the era of behavioral objectives and mastery learning, the distinction is largely overlooked. The student who is correct on 80 percent of the items on a multiple-choice test has "mastered" the objective, without regard to how quickly and easily he performs the task, and without regard to how he might perform under different conditions and different demands. If the objective is fundamental to the learning of another task, the student may come to grief unless he is fluent with the first objective.

Another point can be mentioned only in passing. Assessment is often most meaningful when carried out in a training context (Calfee et al., 1971). Short-term training may serve to clarify the task demands for the student. The teacher can note questions and comments by the students as they perform the task. In the state of California, at least one major assessment project includes a pretest which the teacher is encouraged to give to students until they are thoroughly familiar with how to take the test. Certain commercial tests (e.g., the Stanford Achievement Test Battery) also include short practice tests to familiarize the students with the format and type of content they can expect to encounter. This seems a most sensible practice. More generally, the teacher's assessment should aim to measure the student's response to the ongoing instructional program.

Assessment of transfer

Educators must aim to teach for transfer. Teaching students everything they need to know is impossible. Acquiring knowledge that is transferrable generally requires that the student understand principles as well as basic facts. Transfer sometimes happens automatically, but it is often advisable to teach the principle, and then to check or assess whether the principle actually has been

acquired. Giving many examples of a principle allows students to have experience with a variety of instances where the principle applies. This procedure means that the teacher must be continually checking not only what students have learned, but also whether the student has attained the principles.

How does one assess the extent of transfer? Transfer is assessed by: 1) changing certain features of the situation from those that existed during training, and seeing whether performance remains stable; 2) choosing novel instances of a general principle not part of training, and seeing whether the student can apply the principle; and 3) asking the student to state the principle and supply novel instances exemplifying the principle.

Some years ago, Silberman (1967) demonstrated the importance of assessment of transfer in the evaluation of a beginning reading program. Teaching students to read a list of words by rote is fairly easy; it may be dull for teacher and student, but it can be done. However, when Silberman tested for transfer using a variation on the Esper paradigm, he found that the students had learned what they were taught, nothing more. Using the transfer measure as the standard for a good training program, Silberman proceeded to modify the training program until it worked—until the students learned not only what they were taught, but the principles that allowed them to apply the knowledge in new situations.

Silberman's study focused on decoding skills. For most of us these are so well learned and so automatic that it is hard to remember the lengthy and difficult process of acquisition. Here is a task to refresh your memory: In Figure 10 are six symbol-sound pairs. The symbols—hieroglyphs—were recently discovered on Mars, together with sound tracks. There is every evidence that Martians were a highly analytic and rational folk who would never permit the evolution of a language so erratic and idiosyncratic as spoken or written English. Study these six pairs and then turn to Figure 11 and test yourself against the symbols in the top panel. There are only six pairs, so you should not have any trouble, but many of you will find it rather difficult. Now for the kicker; there are three other symbols in the list, in the bottom panel of Figure 11, similar to the ones you have already seen. If you are smart, you will have figured out the system for this alphabet—it is direct, simple, and rational. Can you name the three bottom symbols?

We can demonstrate the probable existence of individual differences among people. You may have recognized that there is a system to this set of symbol-sound correspondences—if not, you may detect it when you see the system organized in matrix form (Figure 12), where the rows and columns denote critical features. If you have now seen the rule, performance should be error free; indeed, we could readily expand the system. For virtually everyone in this category, the names Chomsky, Halle, and Fant are likely to

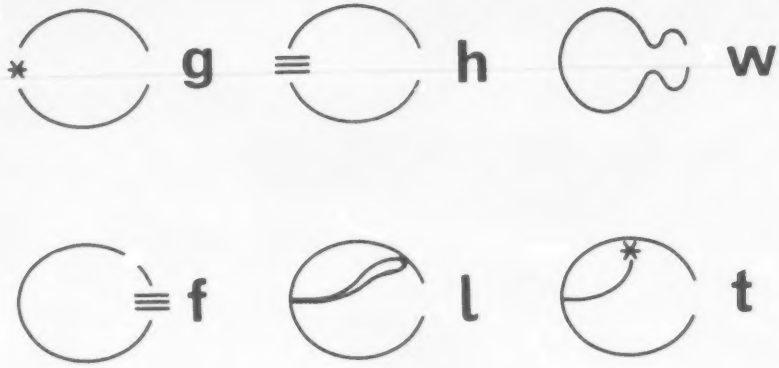


Figure 10. Symbol-sound pairs recently discovered on Mars (after Calfee, 1975).

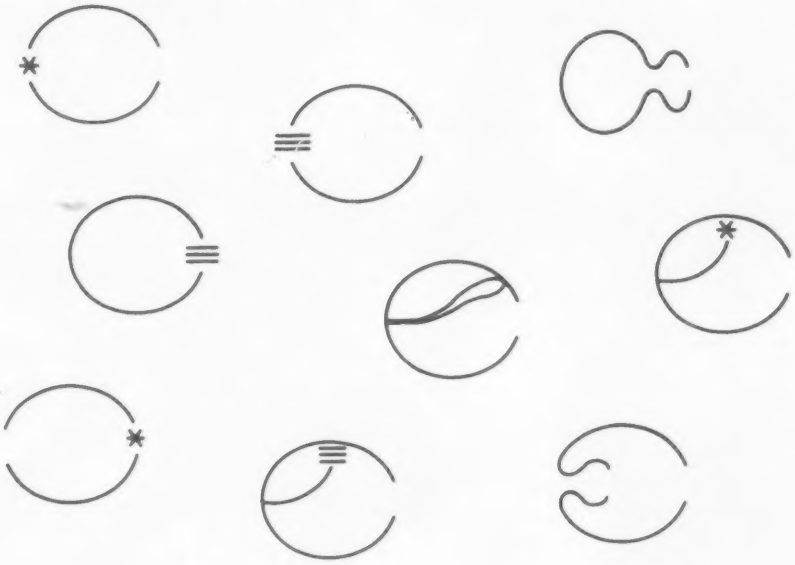


Figure 11. Test item with new items for measuring your recall of Martian symbol-sound pairs.

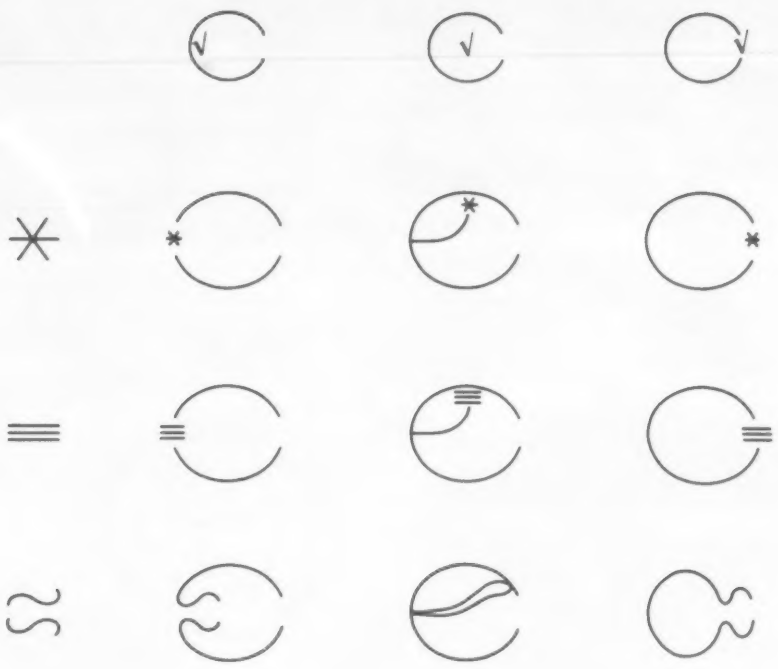


Figure 12. Matrix of Martian symbol-sound pairs.

have special meaning. If our conjecture is right, we could even design a simple readiness test to determine whether an individual would be likely to learn the system quickly.

The system is based on phonological features, rather than on phonemes. One feature denotes the manner of articulation of a consonant (* is stop; ≡ is fricative, ∝ is semivowel), the other feature is the place of articulation (front, middle and back). If you are a phonologist, linguist, or speech clinician, these are meaningful terms, and the writing system no *v* should be readily comprehensible to you. It may not be perfect, but it is reasonably easy to construct and store an algorithm or rule-oriented entry in your memory—easy to construct and easy to use because it is built upon preexisting knowledge. On the other hand, you may not yet know what we are talking about, and then you have to rely on rote memory, which takes time and has limited generality. With enough training, application of the principles become automatic.

Assessment in the classroom

What tests to use

We have made recommendations about the standards for tests designed to serve the classroom teacher in instructional decision making. These recommendations may appear critical of most commercial tests. We are not arguing that standardized achievement tests are "bad", rather we are saying that they serve poorly when used in ways for which they were not intended. But to what is the classroom teacher to turn? Where can he find tests to meet standards appropriate to his purposes? It may be desirable and feasible for publishers and research groups to give some attention to the development and validation of instruments designed for inclass use. However, it is hard for us to imagine a useful system for classroom decision making that is not built in large part around the materials and procedures that comprise the instructional program itself. The classroom teacher may well find the most appropriate materials for assessment within the classroom walls. To be sure, the tests should also provide for the assessment of transfer, i.e., can the student handle related problems that are not part of what has been taught? The idea of asking a student to read aloud from his text seems not to meet current technical requirements of a test. In earlier times, oral reading was a primary assessment procedure. To test for transfer, the teacher might turn to the Gray Oral Reading paragraphs. Teacher training included information on how to make an assessment based on the oral reading and answers to comprehension questions that followed the reading. This still seems a good model for the classroom teacher.

Planning instruction around continuous assessment

Assuming that the teacher seeks to put into practice some of the suggestions given earlier, what are the implications for the rest of the instructional program? We have not thought about this question in any systematic way, but here are a few reflections.

First of all, it becomes especially important to form coherent groups for instruction, to put together those students who share a particular need, who have about the same background to build on, whose needs for the moment are about the same. Re-form the groups as often as possible. When a student changes, pull him out of one group and put him into another. The model here is not three-group reading, but groupings aimed toward specific instructional needs.

Second, develop resources, not just for instruction but for assessment as well. Resources abound in every classroom, in the form of students who already know what somebody else needs to know. A successful tutoring program doesn't happen automatically, but the major dimensions that lead to success are known (Calfee, 1974).

Third, move a student onward to more demanding and difficult tasks only when the foundation is well established, when the student can do the original task both easily and quickly. The student who is pressured to perform a task for which he is not adequately prepared may seek to oblige, but this often leads him to habits and strategies that may work for the moment but lead to negative transfer on later problems. Sometimes the student has to be pulled out of the "regular system" and handled very specially.

Last, the school teacher who accepts a role as an instructional decision-maker will probably have to learn that he must not spend all of his time on direct instruction, in the sense of talking, arranging materials, and the like. It is important to pull back and look where you're going and how you're getting there. Since you're making decisions all the time, it's worth spending some time worrying about these. Ask yourself about the appropriateness of the instructional goals toward which you are aiming. Perhaps you are teaching some students things they don't need to know. Perhaps your goals are at cross purposes with what you really hope to achieve.

In summary

What can the classroom teacher learn about assessment from the research literature? Much and little. The research literature on conventional psychometric testing procedures and instruments serves the teacher in relatively few situations. The teacher should be aware that, for those decisions that entail major and long-term choices for the individual, the program, or the school, a great deal is known about how to acquire the evidence needed for making the choice. While less has been developed to assist the teacher in the day-to-day decisions that occupy his attention most fully, a good model exists—it is to be found in the actions of experimental and educational psychologists as they have sought to understand the intricate workings of the human mind. The guidelines here are fairly clear: be analytic, be precise, resort to controlled variation. The chief need for the teacher may be the development of an intellectual framework that makes it easy to focus on one skill at a time, directs systematic variation in testing conditions, and aids in record keeping. Given such a framework, the teacher can proceed to fill in content appropriate to instruction at the time. The teacher who gains the knowledge necessary to be a good practical experimenter will not need to wait for the results of further research. He will find about him what is needed to carry out the studies that most directly meet his needs—investigations of the cognitive workings of the individual student. For the others, let us hope that soon we will see the systematic research and development that makes it easier for teachers to enhance their skills in assessment, and that provides the tools that facilitate the practice of these skills.

Footnotes

1. Preparation of this paper was supported in part by a grant from the Carnegie Corporation. Many of the ideas are reported in a slightly more technical form in "Assessment for the Classroom Teacher," presented at the Conferences on the Theory and Practice of Beginning Reading Instruction, Pittsburgh LRDC, May 1976. The assistance of Elizabeth Orem, Dorothy Piontkowski, Barbara Tanner, and Barbara Tingey is gratefully acknowledged. A special debt is owed Kathryn Hoover Calfee for her practical counsel about the classroom.

2. In questioning teachers whose classrooms included some kind of compensatory reading program, it was found that less than one in five teachers of sixth grade students made any extensive use of phonics curriculum programs (Rubin et al., 1973). More than 95 percent of the teachers at all grade levels said that comprehension was a major goal. Here is another piece of information from this study that bears on the relative emphasis on decoding skills: In second grade, 75 percent of the teachers report that each child reads aloud to an adult at least once a week. By fourth grade, only 63 percent of the teachers report this much oral reading and by sixth grade, the figure is 57 percent.

References

Anastasi, A. *Psychological Testing*. New York: Macmillan, 1968.

Anderson, R.C. "Can First Graders Learn an Advanced Problem Solving Skill?" *Journal of Educational Psychology*, 56 (1975), 283-294.

Atkinson, R. C., and J. A. Paulson. "An Approach to the Psychology of Instruction," *Psychological Bulletin*, 78 (1972), 49-61.

Barr, R. "The Effect of Instruction on Pupil Reading Strategies," *Reading Research Quarterly*, 10 (1974), 555-582.

Bloom, B.S. *Stability and Change in Human Characteristics*. New York: Wiley, 1964.

Calfee, R.C. "Guidelines for Comprehensive Study of Tutoring in the Elementary School," paper presented under Order No. NIE-P-74-0213. Washington, D.C.: National Institute of Education, U.S. Department of Health, Education and Welfare, 1974.

Calfee, R.C. "Memory and Cognitive Skills in Reading Acquisition," in D. Duane and M. Rawson (Eds.), *Reading, Perception and Language*. Baltimore, Maryland: York Press, 1975.

Calfee, R.C. "Assessment of Independent Reading Skills: Basic Research and Practical Applications," in A.S. Reber and D. Scarborough (Eds.), *Toward a Psychology of Reading*, Proceedings of CUNY Conference. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1976.

Calfee, R.C., R.S. Chapman, and R.L. Venezky. "How a Child Needs to Think to Learn to Read," in L.W. Gregg (Ed.), *Cognition in Learning and Memory*. New York: John Wiley and Sons, 1972.

Calfee, R.C., et al. "Further Explorations of Perceptual and Cognitive Skills Related to Reading Acquisition," paper presented to American Psychological Association Convention, 1971.

Calfee, R.C., and P.A. Drum. "How the Researcher Can Help the Reading Teacher with Classroom Assessment," in L.B. Resnick and P.A. Weaver (Eds.), *Theory and Practice of Early Reading*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, in press.

Calfee, R.C., P.A. Drum, and R.D. Arnold. "Assessment for the Classroom Teacher," paper presented at the Conference on Theory and Practice of Beginning Reading Instruction, Pittsburgh LRDS, May 1976.

Calfee, R.C., and A. Elman, "The Application of Mathematical Learning Theories in Educational Settings: Possibilities and Limitations," in H. Spada and W.

Kempf (Eds.), *Structural Models of Thinking and Learning*. Bern, Switzerland: Hans Huber, in press.
Calfee, R.C., L.W. Fisk, and D. Piontkowski. "On-Off Tests of Cognitive Skills in Reading Acquisition," in M.P. Douglass (Ed.), *Claremont Reading Conference Thirty-Ninth Yearbook*. Claremont, California: Claremont Reading Conference, Claremont Graduate School, 1975.
Calfee, R.C., and J. Floyd. "The Independence of Cognitive Processes: Implications for Curriculum Research," *Cognitive Processes and Science Instruction*. Bern, Switzerland: Hans Huber, 1973.
Calfee, R.C., and K.A. Hoover. "RAMOS Assessment System," unpublished paper, Stanford University, 1974.
Calfee, R.C., and R.L. Venezky. "Component Skills in Beginning Reading," in K.S. Goodman and J.T. Fleming (Eds.), *Psycholinguistics and the Teaching of Reading*. Newark, Delaware: International Reading Association, 1969.
Campbell, O. "Individual Differences in Children's Perception and Comparison of Letter Strings," doctoral dissertation, Stanford University, in preparation.
Carver, R.P. "Two Dimensions of Tests: Psychometric and Edumetric," *American Psychologist*, 29 (1974), 512-518.
Chall, J.S. *Learning to Read: The Great Debate*. New York: McGraw-Hill, 1967.
Cronbach, L.J. "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, 16 (1951), 297-334.
Cronbach, L.J. *Essentials of Psychological Testing* (3rd ed.). New York: Harper and Row, 197[illegible]
Cronbach, L.J., and G.C. Gleser. *Psychological Tests and Personnel Decisions* (2nd ed.). Chicago: University of Illinois Press, 1965.
Green D.B. "The Nature and Use of Criterion-Referenced and Norm-Referenced Achievement Tests," special report by the Association of California School Administrators, Volume 4, Number 3, 1976.
Inhelder, B., and J. Piaget. *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books, 1958.
Kintsch, W. *Learning, Memory, and Conceptual Processes*. New York: Wiley, 1970.
LaBerge, D., and S.J. Samuels. "Toward a Theory of Automatic Information Processing in Reading," *Cognitive Psychology*, 6 (1974), 293-332.
Lambert, W.E. *Language, Psychology, and Culture*. Stanford, California: Stanford University Press, 1972.
Levine, M. "The Academic Achievement Test: Its Historical Context and Social Functions," *American Psychologist*, 31 (1976), 228-238.
McDonald, F.J., and P. Elias. *Beginning Teacher Evaluation Study: Phase II Final Report*. Princeton, New Jersey: Educational Testing Service, 1975.
National Institute of Education. *National Conference on Studies in Reading: Assessment of Reading Comprehension*. Washington, D.C.: NIE, 1975.
Olson, D.R. "A Review of *Toward a Literate Society* edited by John B. Carroll and Jeanne Chall," *Proceedings of the National Academy of Education*, 2 (1975), 109-178.
Rubin, D., et al. *A Descriptive and Analytic Study of Compensatory Reading Programs: Phase I Report*. Contract No. OEC-71-3715. Princeton, New Jersey: Educational Testing Service, 1973
Silberman, H.F. "Experimental Analysis of a Beginning Reading Skill," in J.P. DeCecco (Ed.), *The Psychology of Language, Thought, and Instruction*. New York: Holt, Rinehart and Winston, 1967. Also in *Programmed Instruction*, 3 (1964), 4-8.
Thorndike, R.L. "Dilemmas in Diagnosis," in W.H. MacGinitie (Ed.), *Assessment Problems in Reading*. Newark, Delaware: International Reading Association, 1973.
Venezky, R.L. *Testing in Reading: Assessment and Instructional Decision-Making*. Urbana, Illinois: National Council of Teachers of English, 1974 (a).
Venezky, R.L. "Theoretical and Experimental Bases for Teaching Reading," in

T. Sebeok (Ed.), *Current Trends in Linguistics*, Volume 12. The Hague: Mouton, 1974 (b).
Weintraub, S., et al. "Summary of Investigations Relating to Reading July 1, 1973 to June 30, 1974," *Reading Research Quarterly*, 10 (1974), 267-543.
Weintraub, S., et al. "Summary of Investigations Relating to Reading July 1, 1972 to June 30, 1973," *Reading Research Quarterly*, 9 (1973), 247-513.

Evaluation is an essential part of the teaching process. One way to evaluate the student is to compare the student's performance to some norm. Another way to evaluate is to compare performance with a preestablished criterion. What are Jaap Tuinman's criticisms of behavioral objectives and criterion-referenced evaluation? While critical of criterion referenced measurement, Dr. Tuinman believes that it has an important role to play in education.

Criterion referenced measurement in a norm referenced context

J. JAAP TUINMAN
Simon Fraser University

I wish I could justifiably call this paper, "Beyond Criterion Referenced Measurement," but this would be premature. My views stem from a long term involvement in criterion referenced measurement projects. Analysis of these experiences leads me to believe that unless our profession rapidly moves beyond criterion referenced measurement as it is currently conceptualized and applied, within a short time the very large potential benefits to be derived from the distinction between norm referenced and criterion referenced measurement may well be lost.

In this paper I shall briefly describe a number of interrelated concerns which I'm sure I share with many others involved on a day to day basis with the construction and application of criterion referenced measures in reading education.

The license to measure endlessly

Once upon a time, in the chilly, icy, most southern regions of the globe, a tribe of Penguins decided upon an extensive self-improvement program. A committee of new experts in the art of self-improvement met and decided on the goal of the first Federally Funded SI Program. "We must," they said, "learn to fly." After reading a book by a learned Penguin named Pipham, the committee realized that their global goal needed to be broken down into narrower objectives. "Let's make them behavioral," a farsighted Penguin suggested. "That way we will have fewer measurement problems and less uncertainty about the success of our training programs."

Quickly it became evident that, in order to succeed in formulating adequate objectives, the behavior "flying" needed to be observed first hand. So a subcommittee was commissioned to trek north and observe flying animals in different situations.

One member of the committee made special study of seagulls. He noted that every time just before these birds soared into the air,

they dipped their heads under water. Dutifully he made note of this very important characteristic.

One of his colleagues reported that the birds he observed would peck into the ground and extract a worm just before taking off. The committee decided that the two behaviors showed a great deal of similarity and that, no doubt, both contributed to the mechanics of taking off.

During a month of study, many more analyses were made of the art of flying. The discovery which pleased the committee most, however, occurred when they inadvertently stumbled upon Dallas Airbase and observed the take-off of truly big and heavy birds. "We must analyze their behavior very closely," they told themselves. "Those are the only birds bigger than us." Thus, they added to their list of objectives:

01. "The student must be able to run at great speed in a straight line before taking off."
02. "The student must be able to produce a loud whirring sound from deep inside his/her chest before starting the take-off run."

Satisfied, the Subcommittee on Objectives returned to Penguin Land.

Soon a training program was implemented and criterion referenced measures for each objective were built and validated.

After three months, the first student (a smart Penguin cookie as ever there was) had completed all training modules and passed all criterion referenced tests. She was ready to fly. The Committee on Self-Improvement came out in full force to observe and celebrate the event.

The student's whirring sounded beautiful, her pecking in the hard ice was rhythmic and smooth, her running down the take-off strip an awesome display of athletic prowess. Alas, she never took to the air, and after a full day of frenzied trying, was finally committed to a hospital to recoup from a nervous breakdown.

I suggest that we should not be too hard on the Penguins. Their behavior, relative to our own, wasn't really *that* dumb and unenlightened. In the earlier days of exclusive norm referenced measurement, when a test needed to be built, very stringent constraints on time and length forced test constructors to limit themselves to rather global skills with rather strong face validity. Intuitively and empirically, the relationship between the skills measured and the overall goal—proficiency in reading—was not too tenuous.

In criterion referenced measurement as we know it now, the situation is quite different. To have seventy objectives for one grade level is not too uncommon. Whether they work for a test company or are teachers serving on a summer test construction team, test

constructors basically have a license to add objectives and tests endlessly.

The resulting proliferation of tests is not merely a logistic nuisance; it carries the seeds for a very profound backlash where administrators and teachers alike are getting ready to throw out the entire concept of criterion referenced testing, together with thousands of looseleaf tests.

One of the first orders of priorities is to establish, within the confines of a particular instructional system, which objectives contribute to or are prerequisites for the global goals to be achieved. Eventually, research aimed at settling questions of this kind will contribute to the definition of hierarchies of skills, or objectives to be taught and learned.*

Stennett, Smythe, and Hardy (1975) briefly review the various research strategies for determining which objectives contribute to global goals:

1. Stepwise Multiple Regression
2. Factor Analysis
3. Cluster Analysis
4. Scaling Methods
5. Analysis of Variance
6. Transfer Designs

They justifiably find fault with all of them, yet we know enough to move ahead on this issue. To continue penguin fashion is a far less attractive alternative than to use less than ideal research tools.

McNeil (1974) demonstrated the potential of a very simple technique to identify "false prerequisite skills." His goal was to determine which of fifteen criterion referenced mastery tests in the IOX Battery (which, intuitively, all seem related to success in oral reading) were indeed *not* prerequisites for good oral reading. From a population of oral readers, McNeil took the top and bottom performers. Then he administered to both groups the fifteen IOX tests, including the very face-valid tests as rhyming, affixes, homographs, sounding *r*-controlled vowels, matching sound of short vowels. On the following three tests, 75 percent of the competent readers performed poorly: selecting words with affixes, distinguishing meaning of homographs, and selecting similar sounds of *r*-controlled vowels. McNeil concluded that these tests could not be prerequisites for competent oral reading.

Of course, the McNeil study is a mere beginning. All sorts of methodological questions may be raised in regard to it. Yet the

* I wrote this in 1974. I am far less optimistic now.

pruning must begin. The ultimate benefit of the concept of criterion referenced measurement to reading education is in the balance.

Negating time to learn

Samuels and Edwall (1975) reemphasize that criterion referenced measurement finds its philosophical justification in a model of schooling which asserts that, in the basic skill areas with which we are concerned, most children can learn the content of instruction, provided they are given enough time. Popularly, this view is most often expressed as, "It doesn't matter how often they take the test, or how long it takes, if in the end they show mastery."

I fear that this negation of the "time to learn" variable causes some intricate and basic problems which must be dealt with, rather than suffer benign neglect from all of us. First, if we buy the assumption that criterion referenced measurement finds its ultimate justification in a mastery learning model of schooling, there are two corollaries to using criterion referenced measurement:

1. Instruction *must* provide for an individualized rate of attainment of *en route* objectives.
2. Instruction *must* proceed in a fashion in which no students are forced to deal with next level objectives before attaining mastery in earlier ones.

I will not elaborate on these corollaries, but will merely note that instructional settings in which criterion referenced tests are imposed rarely meet both conditions just stated. A consequence is that such tests are *not* an organic part of an instructional situation but merely a superimposed nicety. In effect, a criterion referenced model of instruction in these cases is superimposed upon a system of instruction which feeds from a norm referenced philosophy.

Many of you probably will react by suspecting me of confusing criterion referenced measurement and testing for mastery. Of course, it is possible to test in such a fashion that scores are interpretable in terms of a specific narrow objective, without reference to other individuals' scores and yet avoid formal reference to a mastery score. Yet, first of all, this is not how criterion referenced tests in most practical applications are used and, second, it becomes doubtful whether any truly meaningful instructional interpretations of such criterion referenced scores are possible.

If the concept of mastery can be given instructional meaning for a particular test and objective, any score interpretation without reference to (non)attainment of mastery can only have diminished and oblique meaning.

A second problem associated with the negation of time to learn is whether it is justifiable, in terms of overall instructional benefits to be gained, to *not* take "time to mastery" in account. Here are some of the questions which must be answered: In which sense

are two students alike, both having reached mastery of Objective A with a 300 percent difference in terms of time needed? How *well* will the mastery of this objective stand up against interference from learning subsequent objectives? Is this a function of rate? How widely will the skill of knowledge acquired in mastering Objective A transfer? Is rate of acquisition a relevant predictor? Are qualitative differences in the mastery of Objective A associated with the rate at which the objective was acquired?

A brief example can clarify the essence of this concern. Roudabush (1974) lists as one of the objectives of the Prescriptive Reading Inventory: The student will be able to identify story elements that set the mood of a story (Objective 77a). Is it possible that Jerry attains mastery quickly because he perceives quickly abstract characteristics of "mood setting story elements" and that Peter laboriously meets the mastery criteria eventually because he relies much more on learning series of specific mood setters?

The basic question to ask is whether the product can be defined without reference to the process or, more specifically, whether the process used influences the relationship of the product to subsequent knowledges and skills to be mastered. I suspect that for many reading objectives beyond the most elementary, the processes utilized to acquire mastery are equally as important for future progress as the mastery itself. Differences in rate, therefore, reflect qualitative differences in processes; they cannot be considered merely losses or gains in time.

My comments here must be sketchy. Speculations as to which variables may underlie rate differences for specific objectives must be foregone for the moment. As a group, they probably may be referred to as general intellectual functions.

If one acknowledges that rate of learning is a function to be measured even within the context of criterion reference measurement, two strategies suggest themselves. The first one is widely implemented without, it appears, much reference to underlying theoretical issues. This strategy calls for *de facto* limits on the time available for attainment of certain objectives. All students receive the same tests at a particular time. For many students, failure to demonstrate mastery in this case reflects a slow rate of acquisition.

We have just recently started work on the feasibility and desirability of having criterion referenced tests result in two scores. The first score, the mastery score, is obtained from a subset of items for which all students have sufficient time to finish. Failure to reach criterion cannot be a function of a slow working rate. The second score, the efficiency score, is in effect a speed score. Thus, we are deliberately reviving the old power versus speed issue from norm referenced tests within the context of criterion referenced testing. Efficiency is of particular interest in those tests which involve application and integration of lower order skills. Anticipating later

comments about the utility of criterion referenced skill scores in estimating scores on norm referenced tests, it may well be that CR mastery scores differ from efficiency scores in predictive power.

Comparisons in disguise

Apparently, the most appealing feature of CR measurement is the claim that, "A child is not judged on the basis of peer performance, but he is judged against an objective criterion." This concept never fails to thrill students in measurement classes.

The question is whether the onset of CR measurement in our educational practice actually has made a difference in this respect. Can it? Should it? Is *this* the major *real* difference between criterion referenced and norm referenced measurement?

Norm referenced measurement has come to mean that scores of an individual are interpreted on the basis of peer performance *as forced* into a normal curve. The last qualifier is important. Half of the objections against the practice of norm referenced measurement probably are levelled only against the fact that postinstructional performance often is not normally distributed. It is, however, quite possible to conceive of norm referenced measurement using other curves. The notion of determining performance curves of students who have been taught under optimal instructional conditions is especially intriguing to me.

Let us assume that a group of 100 students is taught to Objective A under optimal circumstances. Whether this has led to complete mastery for all 100 students is indeterminable. All evidence is contained in a set of test scores heavily skewed to the right. Let us say that Sarah scores 7 out of a possible 10 with 8.5 constituting mastery, by definition. Let us also assume that 8.5 is the mean score obtained. The following statements about Sarah's score are meaningful:

a. She scored below mastery.
b. She scored below the mean score which was at the mastery level.
c. She scored below 90 percent of the students, 85 percent of whom scored at or above mastery.

I consider observation (c) more meaningful than (a) or (b). Why? It tells me not only that Sarah has not reached mastery, but that her performance is rather unusual. Remember that in the peer comparison, no predetermined curve was forced onto the data.

If, indeed, the type of criteria used in educational measurement were intrinsically meaningful, we possibly would be in a position to avoid any reference to peer group performance. However, criteria on reading tests seldom can be defended with reference to reading only. We must invoke "the mature reader," "the average reader," "the functionally literate reader" in order to defend our

selection of relevant behaviors. To make matters worse, criteria in terms of number of items right on a specific test still are almost entirely a matter of convention and intuition.

In fact, what is typically happening is a process of delayed norm referencing. In brief, this is the sequence of events: An objective is selected. Items are written. A level of performance for mastery (80 percent, 90 percent) is specified. Then the test is used. In favorable circumstances, the first usage is an experimental tryout. When the results come back and neary all students (after instruction, of course) reach criterion, the test is reproduced on full scale. If, however, it turns out that too many students do *not* reach mastery, items are revised until the test is easy enough. To "pass" the test from here on out is a norm referenced achievement. A performance is successful only because of the relative (in)ability of a specified peer group.

The process used is, in essence, no different than the time-honored procedure of building a norm referenced test. Only the parameters of the distribution of scores have changed.

Can this situation be remedied? Are true criterion referenced tests in reading possible? Theoretically yes—pragmatically speaking, possibly not. There are some reasons for this pessimistic outlook.

Specification of objectives at the elementary level involves two dimensions: the perceptual or cognitive activity demanded and the language elements (words, sentences) which form the population of elements with which these activities are realized. There is no practical way to calibrate language or define relevant characteristics so that items of comparable difficulty and complexity can be generated from them.

Example: The student will be able to identify the correct definition of a word as it is used in a given sentence.

The operation demanded here is relatively straightforward: Identify the correct definition. The problem lies in the multitude of choices available in terms of key words *and* sentence context formation. For instance, even if a third grade word list (in itself usually a somewhat ambiguous item) is used, the item writer has few a priori means of judging the fit of his item to an ideal criterion.

Consider the following sentences:

1. I saw a ratmilli blooming in the desert.
2. When I traveled through the desert, I saw a ratmilli.
3. Did you give a ratmilli to mother?

A ratmilli is a:

a. present
b. flower
c. desert fox

Notice how every sentence works differently in terms of providing or not providing help for identifying the correct alternative.

In norm referenced measurement, with an a priori established ideal distribution of scores, this kind of ambiguity is not very bothersome. In which sense, however, can a criterion be meaningful if its meaning cannot be specified in reference to the operation performed to create the items measuring it? For contrast, compare the measurement problem associated with the following objective: The student will be able to reassemble a completely disassembled Volkswagon carburetor of which the parts have been arranged in random order. No variance exists here in the elements upon which the task called for is performed from test to test.

The only avenue available to escape the dilemma of meaningful criteria in criterion referenced tests is acceptance of the concept of *critical* criteria as for instance are used in selection theory. Here a criterion is formulated in terms of the probability of success on a future task by those who reach criterion and the probability of failure by those who don't.

Thus, in terms of a test for Objective A, the question becomes: Which score is associated with a too high probability of failure on the next objective? Notice that this interpretation of mastery assumes a hierarchical ordering of objectives and that some means for defining "risk" is presupposed. Much of the research on hierarchical relationships of objectives (which I called for earlier) would result in information which might be used to establish meaningful criteria scores.

Presumably, the particular form of instruction chosen to lead students to mastery of a set of objectives will influence the relationships among them. This is no small obstacle to establishing meaningful criteria. For this reason, though theoretically the potential for meaningful criterion referenced measurement exists, I don't see an immediate practical solution to the problems.

Do I advocate the rejection of criterion referenced measurement in reading? By no means. The philosophy underlying CR measurement is sound. Why do children have to learn more than is needed? (I am aware of the "needed for what" implications.) Why do they have to beat out another student to prove adequacy? Why do we insist that postinstructional performance be treated statistically as raw psychological variables? Therefore, over the next few years, we must be very careful in how CR measurement is handled in educational practice. No amount of massive data juggling, as already is going on in regard to CR measurement, will make up for lack of a clear conceptualization of our problems.

Here, by way of summaries, are some theses.

1. Criterion referenced measurement must be conceived of both in terms of Skills Management Systems (e.g., Wisconsin Design) and in terms of global achievement.

2. In the context of skills management, the following issues arise:
   a. Hierarchical relations among objectives—empirical solutions are possible but tenuous.
   b. Establishing critical criteria—empirical solutions are possible, provided someone makes a clear judgment about what constitutes "risk."
   c. The issue of how the objectives are attained, through direct instruction or indirectly by teaching methods which do not require fractioning of skills *during acquisition*, is independent of whether and how these skills should be measured.
   d. The issue of rate to learn and rate of integration—empirical determination of the payoff of efficiency scores is possible.
3. In the context of CRM as an evaluation of global performance, these issues arise:
   a. Adequacy versus mastery. Does one *master* reading or does one surpass a minimal level of *adequacy*? I opt for the latter. This means that adequacy scores must be developed either consensually or via the critical criterion route, tying in reading performance with later school or job success. We virtually know nothing about this.
   b. We cannot escape global judgments. Per item testing and reporting is convenient. The interpretation problems are baffling, however. What does the National Assessment of Educational Progress data mean, for instance?
   c. In a sense, standardized *global* criterion referenced measures should replace the norm referenced variety. Curricular content validity becomes secondary to criterion (not criterion related) validity.

References

McNeil, J. O. "False Prerequisites in the Teaching of Reading," *Journal of Reading Behavior*, 1975.

Roudabush, G.E. "An Empirical Structure for Reading Objectives," *Journal of Reading Behavior*, 6 (1974), 403-420.

Samuels, S. J., and G. E. Edwall. "Measuring Reading Achievement: A Case for Criterion Referenced Testing and Accountability," *Measurement in Education*, 6 (1975).

Stennett, R. G., P. C. Smythe, and M. Hardy. "Hierarchical Organization of Reading Subskills: Statistical Approaches," *Journal of Reading Behavior*, 1975.

Whose responsibility is it to improve the method and materials used in the teaching of reading? Professional types who do research in the psychology of reading? Curriculum specialists? The developers and publishers of reading textbooks? The classroom teacher? The reading teacher?

All of the above persons will find themselves in the following chapter, and each one will see that he has an important role to play in the process of improving the teaching of reading. By identifying what his role has been in that process, Wayne Otto hopes to influence others to become similarly involved.

Less concerned with what educational researchers have to say to the classroom teacher, this author would rather see the lines of communication opened so that all involved parties—researchers, developers, and teachers—will begin to share with one another what they know about reading and what they still need to find out.

A pragmatic-empirical approach to research in reading

WAYNE OTTO
University of Wisconsin at Madison

Research in reading ought to be a two-way proposition; teachers of reading probably have as much to say to researchers about research as research has to say to teachers. That is the main thesis of this paper. And the pragmatic-empirical approach is offered as one way to unite researchers, teachers, and developers on the total team, in their quest for improved teaching of reading. The term *pragmatic* implies to me that one's grasp ought to be about equal to one's reach, and I am inclined to say, "If it works, do it." I also believe in empiricism. By this I mean that one should collect data and observe rather than engage in armchair speculation. My belief in these two principles makes me a pragmatic-empiricist.

The research-development dilemma

David R. Krathwohl titled his 1974 Presidential Address to the Division of Educational Psychology of the American Psychological Association, "Perceived Ineffectiveness of Educational Research." [The quotations used in this paper are from a revised version published in the *Educational Psychologist*, 11, (1974), 73-86.] The pragmatist in me wants to say that "perceived" could have been dropped, making the title a stark "Ineffectiveness of Educational Research." My title would NOT have described Professor Krathwohl's message. But it most assuredly would have summed up a conclusion that I—with extreme reluctance—have come to: The preponderance of what is and has been passing for and published as educational research is ineffectual. Of course, we can identify certain research that has had profound effects on both educational theory and educational practice. But the preponderance tends to be pure twaddle and I think almost everybody who thinks about it knows it. In our field, much of what is published as research

related to reading is related to reading only in the most obtuse way and has no real importance to anyone—except perhaps the researchers.

Krathwohl was trying to deal constructively with a problem. He spoke about one aspect of it as follows:

> There is continual pressure to shorten the intial stages of [the research and development] process which is resisted by the true empiricist. He is troubled by missing and/or too markedly abbreviating the confirmatory steps, for he fears users will base their actions on research which later will invalidated. Indeed, even if no harm were done to the consumer by use of a product later found to be invalid, such an occurrence gives research a bad image. Potentially it could sour the public on support of research. For similar reasons, he may be equally dismayed when work in incomplete stages of development is put to use and fails. It then reflects badly on the developers and the processes they used (p. 78).

The research-development dilemma is well known to anyone who has ever tried to engage in that process. The process looks so beautifully simple and logical when it is described on paper. But make your move too soon and you become the object of ridicule and scorn when what seemed to work perfectly in its pristine state falls flat in a real world tryout. Make it too late and you find yourself talking to an empty auditorium.

Krathwohl pointed out another aspect of the same problem:

> ...the process [of research and development] is a very long one for practitioners faced with the daily press of problems that research is supposed to help alleviate. It operates in a time frame which is a world apart from the one in which he must operate. And the fact that the process sometimes is extended because the developer will not release a product until it is engineered into a "teacher proof" package makes the user doubly resentful, first because of the period the product is withheld from use and, second, because of his implied incompetence which results in the need to "teacher proof" products (p. 78).

So there you have the problem that has been experienced by many teachers who have gone skipping down the yellow brick road of research and development. The road turns out to be long, it gets rocky, and it is beset by witches. If we can squeeze that metaphor a little bit more, we need to figure out ways to get to see the Wonderful Wizard of Oz.

Krathwohl suggested several ways to improve the outcomes of the educational research enterprise. Two of them seem most relevant to resolving the research-development dilemma identified here. Both have to do with bridging the gap between researcher and user, *especially* the practitioner teacher.

First, researchers can begin to think of themselves not so much as producers of knowledge as producers of findings to be confirmed in practice. Speaking for the researchers, Krathwohl says:

> ...rather than presenting our findings as proven, our presentations should make clear the tentative nature of our findings and routinely seek the help of practitioners in further validating or invalidating these findings, suggesting the simplest possible methods for doing so (p. 83).

The big change is that instead of being content merely to point out "implications for further research" to others of their own ilk, researchers would be inviting teachers to participate in testing the validity of their findings. The change is indeed a big one. Ultimately, it could lead us to ask a researcher, "What impact has your research had on practice?" and not, "How many articles have you published?" Equally important, it could also establish a solid base for the validation efforts of practitioners. It's time that teachers stopped flitting endlessly from one innovative flower to the next without even thinking about the possibility of changing the flowers.

Second, developers can consider modifying their model, particularly with regard to the extended time lines and production of "teacher proof" products. Krathwohl put it this way:

> ...developers may wish to research the most appropriate level of development relative to "teacher proofness" in terms of kinds of teachers, schools, and instructional goals. If materials can be put in the field at earlier stages of development, this would both reduce the time and expense of development, as well as give teachers more of the creative and intellectual challenge which many of them want as growing professionals (p. 84).

Or better still, developers can consider teachers full partners from the outset and focus as much or more on the *process* as on the *product* of development.

How can we—researchers, developers, teachers—get to see the Wizard? The answer is simple: We all skip down the yellow brick road together. But it isn't easy because getting to *see* the Wizard is the hardest part. It takes common goals, collaboration, mutual respect, and common sense.

Pragmatic-empiricism in practice

A pragmatic-empirical approach to research in reading offers a way to deal with the research development practice dilemma. It offers us a chance to see the Wizard. Rather than tell you what you should do, let me tell you what I do to put the approach into practice.

The good news and the bad news

In the first place, it should be clear that most of my current work has to do with development. I find myself in between the hard line researcher on the one hand and the soft line practitioner on the other. Let me explain. Most of the people whom I consider real "researchers" are what I would call hard liners: If it comes to choosing between elegant design and rigorous methodology on the right hand and a commonsense approach to a practical problem on the left, they'll *always* go for the right. The good news is that their results are impeccable in terms of clarity and probability. The bad news is that they tend to tell us more and more about less and less. Most of the people I consider *real* practitioners—teachers—are what I would call soft liners: If it comes to making a choice between a practice or product that has known flaws but could be improved and possibly greatly enhance children's learning on the right hand and a practice or product that nobody really understands but seems to work okay on the left, they'll *usually* go for the left. The good news is that they keep school. The bad news is that it never gets any better.

What's needed, of course, is what lots of people for a long time have been saying is needed: Something or somebody to bridge the gap between the hard line researchers and the soft line practitioners. We attempt to bridge the gap with a pragmatic-empirical approach.

The empirical tryout and revision cycle

This is what we do. We listen to the practitioners tell us what they can do and what they need. We listen to the researchers tell us what they know and what they think they might be able to do. Next we pick out one of the practitioners' needs and come up with one commonsense alternative—one that reflects all the research results we can handle—for responding to it. For a while, we put all our eggs in one basket and develop the alternative we have chosen—with continuous input from practitioners and input as needed from researchers—until it is as good as we can get it. And then we try it out, first on a very limited scale with lots of support. If it seems to work, we try to find out why because, ultimately, it has to work without more support from the developers. If it bogs down, we try to find out why. If we find a problem, we revise. If we see a gap, we add on. We do a hard line study when we think we know enough to plan one. We do a soft line tryout when we need to feel our way. If we're dealing with a product that's used in the classroom, we look at it in at least two ways. We find out how *teachers* perceive it, whether they are using it as we thought they would, and whether they want to continue. And we find out how kids perceive it, whether they can handle it as it comes to them, and whether it makes any difference in terms of achievement or anything else.

Some problems

The following list reflects some of the factors that in practice complicate the above approach.

1. Note that the approach requires the collaborative efforts of—or at least substantial input from— researchers, developers, and teachers. Try to do it all yourself and you're spread too thin to get anything significant accomplished. Cut out or soft pedal input from any one of the sources and you wind up with an end result that a) doesn't make sense, b) doesn't work, c) nobody wants, or d) all of the above. The hard part, of course, is to strike a working balance.

2. The working balance is important. The hard line researchers always insist that they need at least two more decades before they will be ready to release their findings. Teachers insist that they need the product tomorrow because the Board of Education wants *everybody* to read at grade level by next week. The developer needs to come up with something that a) has the unequivocal backing of research and theory; b) is consistent with current trends in curriculum development; c) is universally appealing to teachers, administrators, children, the public, and funding agencies; d) shows no bias with regard to race, religion, sex, handicapping conditions, or anything else; e) is not harmful to the environment or to individuals, even when taken internally; f) can be packaged neatly; g) doesn't cost much; and h) will be attractive to a publisher. The balance can be struck, but everybody has to give a little. And common sense helps a lot.

3. The process is time consuming, so there must be a continuing commitment from all concerned. If we start with what we know in terms of research results and what we need in terms of teachers' perceptions, we know from the start that some excursions are likely to be needed: additional research to clarify a point, corrections when data fail to yield positive results. Participation in such a process is difficult and sometimes discouraging, particularly with the on-again, off-again funding we've come to know in recent times. But the alternatives can't be very appealing: either making do with what we've got, with no attention paid to trying to improve it, or flitting forever from one alternative to the next in an endless quest for the best.

4. If time is important, *timing* is equally so. Some reasonable balance must be struck between the empiricists' demands for rigor on the one hand and the pragmatists' demands for action on the other. One thing we've done is to put out a "developmental edition" of certain products. That means a product has had limited tryout but we recognize that there are still basic problems and major revisions are continuing. At the same time, we feel that the product is as good or better than the available alternatives. It is disconcerting to some teachers to know that a product will change substantially in the

relatively near future, but the alternatives are either to keep a product that is already "workable" and reasonably effective on the shelf or to stop development short of an attainable goal.

The list could go on, but the point is that the pragmatic-empirical approach must be implemented with a pragmatic-empirical attitude.

How to change the flowers

I've described a developer's use of a pragmatic-empirical approach to bridging the gap between research and practice. The process requires getting researchers, developers, and teachers into the act of improving the teaching of reading. Teachers are *involved*. But how can teachers take the incentive in getting researchers and developers to tackle the problems that *teachers* think are important?

My observation is that, too often, teachers are inclined to either expend great energy in "trying out" innovation after innovation in their quest for the best approach to teaching reading, or be so turned off by the helter skelter of innovation hopping that they expend equally great effort defending and clinging to a status quo they know could be improved. Too seldom do teachers direct their efforts toward becoming personally and productively involved in a process that starts with the identification of classroom problems and ends with improved classroom practices. But that's what it takes to change the flowers.

Teachers may remain passive participants in reading research and development. It's a comfortable trip with some rewards and no real responsibilities. But you are not satisfied or you wouldn't be reading this. You want to change the flowers because you have some fair notions about what's wrong with them now and you know how you'd *like them to be*. If you have already taken the first step of joining a researcher-developer-teacher team, think about taking the next step. Start your own team. Then you can tackle the problems and seek the results that you know are important in the classroom.

The ultimate reality of teaching reading remains in the classroom. The teacher can assure the *pragmatic* in a pragmatic-empirical approach to research in reading.